VOL. X 1919

THE JOURNAL
OF
Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

EDITORS:

W. C. BAGLEY,
Columbia University.

C. E. SEASHORE,
University of Iowa.

J. CARLETON BELL, Managing Editor.
Brooklyn Training School for Teachers.

GUY MONTROSE WHIPPLE
Carnegie Institute of Technology

BALTIMORE, MD.
WARWICK & YORK, INC.
1919

TABLE OF CONTENTS

INDEX TO NAMES

Authors of original articles, communications or discussions have names printed in CAPITALS, and the page numbers in **bold face** type. Authors reviewed, contributors of reviews, and authors mentioned have names in Roman, and page numbers in light face type.

INDEX TO SUBJECTS

VOLUME X JANUARY, 1919 NUMBER 1

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE IMPORTANCE OF DIAGNOSIS IN EDUCATIONAL MEASUREMENT

ARTHUR W. KALLOM
Department of Educational Investigation and Measurement, Boston, Massachusetts

THE GROWTH OF EDUCATIONAL MEASUREMENT

Since the movement for educational measurement first started, there has been a great increase in the literature and in the number of tests which are available for measuring educational results in the elementary schools. The movement began through the demand for the educational survey. This survey was made by experts in education from outside the city system. It was the intention to thus obtain an impartial and just statement of the conditions as they existed in the city which was being surveyed. The movement has spread until by 1917 the United States Commissioner of Education has reported over ninety different surveys. Also there are a large number of cities which employ their own director,[1] whose work consists in giving standard tests from time to time and measuring the results of the educational product in the city in the same unprejudiced and impartial manner which is obtained by having the outside expert come in to do the testing, with the advantage that the director becomes familiar with the city, its civic and educational policies, and may thus make his plans fit in with other movements.

The school surveys have shown to the educational profession the lack of three important elements in our educational administration.[2]

(a) Lack of standards of achievement
(b) Lack of standard tests
(c) Lack of a record of educational experience

Lack of any standard achievement is shown by the uncertainty in the mind of the teacher of what ought to be expected from any pupil in the school system. That is, how many examples of a cer-

[1]Educational Standards, vol. V, May, 1915, Boston.
Ballon, F. W., [2]School and Society, vol. V, No. 108, pages 61 to 70.

tain kind should be performed in addition within a given time? What per cent. of accuracy should be expected of a pupil in the seventh grade in subtracting examples of a certain kind within a definite period of time? How much of the geography which was taught in the fifth grade should be retained by pupils in the eighth grade? The answers to all of these questions and others like them are very indefinite in the teacher's mind.

This lack of any standards of achievement is directly dependent upon the lack of standard tests. The ordinary examination given by a teacher is not a standard test because the teacher does not know the elements of which the test consists. That is, she does not know the relative value of the different questions of which the test is composed. Even if the test partakes of purely mechanical matter as the adding of columns of figures, the teacher does not know whether all combinations have been included, nor does she take into consideration the relative difficulty of the different combinations. So it has been only since the inception of educational measurement, that standard tests have been available and even now tests along certain lines are difficult to prepare.

The lack of any record of educational experience is nearly, if not quite, as great as the lack in the other two respects. Many of our good teachers, however, have kept a record of their experience in various subjects from time to time. For example, some teachers have kept a list of the words which have been missed by children in their composition work and have used these words as a basis of their spelling lessons. They have also kept more or less of a record in connection with their arithmetic. These, however, have been incomplete and have not been systematized. Neither has this material been organized and passed on so that others may profit by the mistakes which have been made. In general the more complex a subject or topic, the less detailed is the record which has been kept.

Important Phases of Educational Measurement

The inauguration of any department of educational measurement in a public school system is only defensible on the grounds that it has for its ultimate end the better education of the individual pupil. All departments of the school administration including superintendent, supervisors, principals, and teachers are for the sole pur-

pose of educating the pupil in the most efficient manner possible. The maintenance of any one of these departments can only be justified if it performs this particular service.

Standards of Achievement

How may educational measurement help the teacher in performing this service of educating the individual boy or girl? The first work of any department in educational investigation is to set up a definite standard of achievement in the different branches of the curriculum as fast as possible. This standard should be based not on any theoretical basis, but upon the actual work of the pupils taking a standard test under supervised conditions. It is not possible to discuss in this paper what those conditions may be. Suffice it to say at this time, that those conditions must be such that the person in charge of the test knows what they are.

Record Cards Showing Achievement

This standard now becomes a goal which may guide the teacher in the instruction of the children. Her effort and energy may now be directed toward those pupils who have not reached the standard of achievement as set by the city instead of having her work distributed over a class, *part of which does not need the extra work which is being given.* These individuals are pointed out because the teacher keeps a record of each test on a card which is uniform for the city and is sent with the pupils as they go from class to class in the same school or from school to school within the city.

After this standard of achievement has been found, the pupil also has a knowledge of what he is expected to do. He knows whether his achievement is above or below the standard set for the city. If he is below, he becomes anxious immediately to raise his attainment and come nearer to the standard as set for the city. If he is above the standard, he then immediately becomes anxious to maintain his record. Thus the setting of a standard of achievement on a city-wide basis becomes a very important motive in his work.

Figures 1, 2, and 3 show the records of three individuals in work with Courtis tests, series B, with the explanation, diagnosis, and treatment in each case. The solid line in each figure indicates the standard, while the dotted line indicates the score of the individual.[1]

[1] Taken from School Document No. 22, Boston Public Schools, "*The Value to the Teacher, to the Principal, and to the Superintendent of Individual and Class Records from Standard Tests.*"

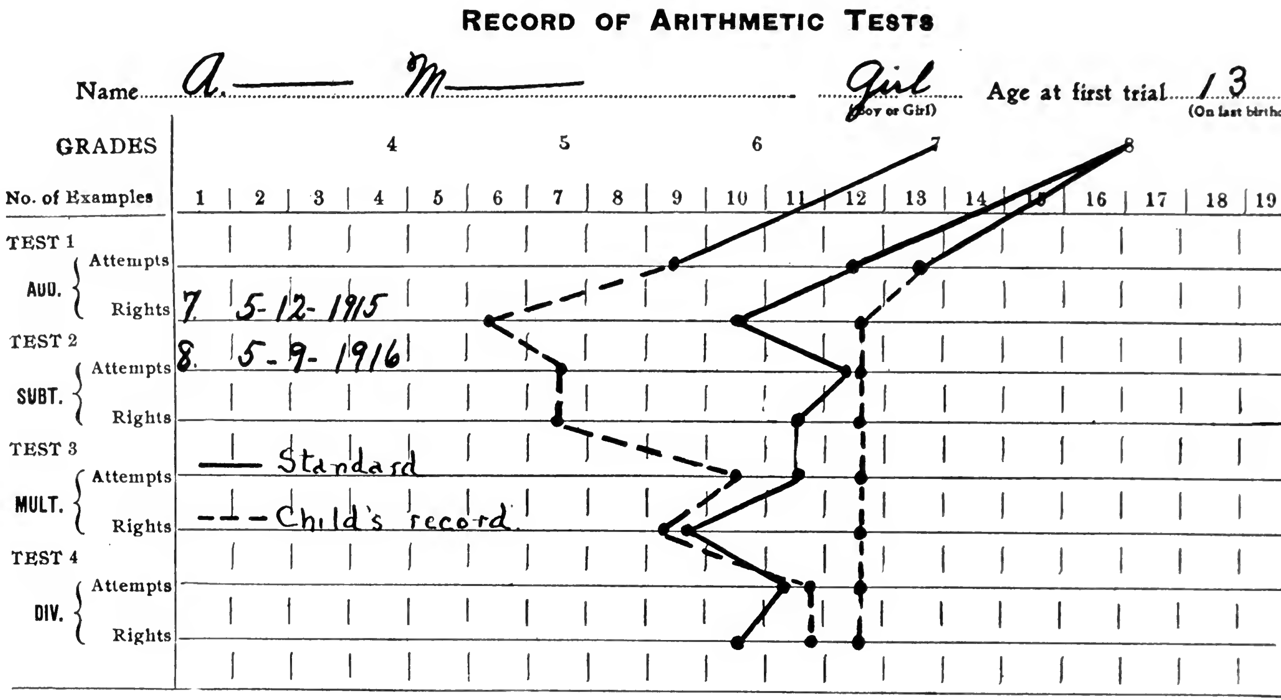

RECORD OF ARITHMETIC TESTS
Name A.—— M——
Girl
(Boy or Girl)
Age at first trial 13
(On last birthday)
GRADES
4
5
6
7
8
No. of Examples
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
TEST 1
AUD.
Attempts
Rights
TEST 2
SUBT.
Attempts
Rights
TEST 3
MULT.
Attempts
Rights
TEST 4
DIV.
Attempts
Rights
7. 5-12-1915
8. 5-9-1916
Standard
Child's record

FIG. 1

Interpretation of Fig. 1

EXPLANATION.—This shows the graph of a girl graduated from an elementary school in June, 1916, who was thirteen years of age when she first took the test. Her record shows two tests: May 12, 1915, when she was in Grade VII, and May 9, 1916, when she was in Grade VIII.

DIAGNOSIS.—When she first took the test she showed uneven ability in the processes. She was slow in addition as well as inaccurate. She was accurate in subtraction but slow. Her multiplication score was up to eighth-grade standard in the number of examples correct, while the division score was up to the eighth-grade standard in speed and showed perfect accuracy. In the second test she showed remarkable uniformity in the four processes, attempting twelve examples in every process but addition, in which she attempted thirteen, and in every case getting twelve examples correct.

TREATMENT.—The second score is an unusual one and shows an ideal toward which to work. While the number of examples attempted is at or near the standard, there is at the same time almost perfect accuracy. Just enough emphasis has been placed on the deficiencies in addition and subtraction to eradicate them and bring ability in those processes up to the ability in multiplication and division. The record of this girl is the kind of standard to which the department looks forward. It would constitute a better standard than the present one. When pupils enough to make a change reasonable have succeeded in achieving a result such as this girl's, the present standard will be modified accordingly.

RECORD OF ARITHMETIC TESTS

Name A —— B ——, Girl (Boy or Girl) Age at first trial 11 (On last birthday)

School District .... School .... City Boston

DATE	GRADE		SERIES B TEST 1 ADD. Attempts	TEST 1 ADD. Rights	TEST 2 SUBT. Attempts	TEST 2 SUBT. Rights	TEST 3 MULT. Attempts	TEST 3 MULT. Rights	TEST 4 DIV. Attempts	TEST 4 DIV. Rights
1-8-1914	VI	1st Trial	7	2	5	1	5	4	4	0
4-21-1914	VI	2nd Trial	8	5	5	1	5	3	3	0
10-9-1914	VII	3rd Trial	9	5	5	2	8	6	4	2
5-17-1915	VII	4th Trial	12	12	11	2	11	11	6	0
5-11-1916	VIII	5th Trial	17	17	20	14	14	12	19	19
		6th Trial								
		7th Trial								

FIG. 2

Interpretation of Fig. 2

EXPLANATION.—A——B——graduated from an elementary school in May, 1916. She took the test five times; on January 1, 1914, and April 21, 1914, when she was in the sixth grade, on October 9, 1914, and May 17, 1915, when she was in the seventh grade, and on May 11, 1916, when she was in the eighth grade.

DIAGNOSIS.—The addition score improves at every trial. From the score of 7–2 it finally goes to 17–17, perfect accuracy. In subtraction the increase in the number of examples attempted is just as marked; the growth in accuracy is not so noticeable, although it is evident that the accuracy rises from 20 per cent (5 attempted, 1 right) to 70 per cent (20 attempted, 14 right). The multiplication score, which shows good accuracy in the first test, although slow in attempts, rises to 14 examples attempted with 12 right. The division score in the first and second tests shows that the girl attempts only 3 or 4 examples and is unable to get 1 right. In the third test she finds herself able to perform correctly 2 examples out of a total of 4 attempted. In the next test she attempts more, but lapses to none right. Between that and the fifth test she evidently learns how to divide, as shown by her next score of 19–19.

TREATMENT.—This girl was exceptionally inaccurate in every process except multiplication. Drill has been applied so that noteworthy progress has been made in those processes which needed it most. In addition and division progress is especially noticeable. The deficiencies have been eradicated. There has been perhaps a little over-emphasis as the scores exceed the standard. For example, in 1915 the addition and multiplication scores had reached the eigh-grade standard. Therefore, this child needed only enough drill on those processes to keep the record at the standard.

RECORD OF ARITHMETIC TESTS.

Name I—— L—, Boy (Boy or Girl) Age at first trial 13 (On last birthday)

School District .... School .... City Boston

DATE	GRADE		SERIES B							
			TEST 1 ADD.		TEST 2 SUBT.		TEST 3 MULT.		TEST 4 DIV.	
			Attempts	Rights	Attempts	Rights	Attempts	Rights	Attempts	Rights
1914-1-8	7	1st Trial	8	5	9	5	5	4	5	0
1914-4-23	7	2nd Trial	10	2	20	13	13	10	7	3
1915-5-12	7	3rd Trial	14	6	13	12	11	6	8	6
1916-5-9	8	4th Trial	13	7	18	15	14	8	9	2
		5th Trial								
		6th Trial								
		7th Trial								

FIG 3

Interpretation of Fig. 3

EXPLANATION.—This is the record of a boy who graduated at the age of sixteen from an elementary school. He spent two years in Grade VII and three trials of the test are recorded in that grade. The other trial recorded is when he was a pupil in Grade VIII.

DIAGNOSIS.—By an examination of this boy's record it will be seen that in four successive tests he made little or no improvement in his addition and division scores, even though they were shown to be low in the first test. There is variability also in the subtraction and multiplication scores with some improvement in speed.

TREATMENT.—This report has proved of no value to the teacher as she did not use it to diagnose the pupil's difficulty and prescribe for him the particular kind of drill he needed. It illustrates well the futility of class drill in eliminating individual difficulties. The child goes blundering on, meeting the same difficulty and making the same kind of mistake. The record of the first test showed especially that something serious was wrong with this boy's ability in addition and division. Yet his teacher did not make proper use of that record and the inability showed itself again in the next test. His next teacher did not use the record and it went on until this boy graduated from an elementary school with the ability to perform correctly only two long division examples in eight minutes and only seven addition examples in the same length of time.

Use of Record of Arithmetic Tests

Such records give the information necessary to know which pupils need individual help. By means of these cards the teacher is able to determine immediately upon the entrance of a class into her room what pupils are poor in any given branch of the work which has been tested. This not only saves time ordinarily used for finding out these facts, but insures that the knowledge which the teacher has is much more dependable than any she might gather for herself. Not only does she find out from these records that a particular pupil may be poor in fundamentals, but she finds in what particular operation the pupil is poor. For example, the pupil may be relatively poor in arithmetic, but his record may

show that his poor work is determined largely by his work in addition and multiplication, whereas in subtraction and division his work may be exceedingly good.

Urgent Need for Methods of Diagnosis

The knowledge, however, that a particular pupil is weak in addition does not necessarily tell the teacher in what particular phase of addition the individual may be weak. The results obtained in such a standard test as the Courtis standard tests in arithmetic, simply indicate a condition and do not in themselves diagnose this condition. A low score is a symptom; the interpretation of the sympton is quite another matter. For example, there are a number of different factors entering into the addition of integers. A pupil weak in any factor would necessarily show a poor record in adding examples consisting of nine numbers of three digits each. Some of the more important factors in the operation of addition are as follows:

1. Simple addition combinations
2. Single column addition of three figures each
3. "Bridging the tens" as 38 plus 7
4. Column addition, seven figures
5. Carrying
6. Column addition with increased attention span, thirteen figures to the column
7. Addition of numbers consisting of one or more digits

Thus it will be seen that a pupil may be weak in his ability to carry from one column to the next and, therefore, be weak in his ability to do addition examples. On the other hand, he may be able to carry correctly but his span of attention may not be sufficiently strong to enable him to add columns of eight or nine figures and for this reason he would show a poor result in his ability to add integers.

When the teacher has found out the fact that a pupil is weak, it becomes necessary for her to use her resourcefulness to determine at what stage the pupil is weak. This is made comparatively easy by the best forms of practice material which are on the market at the present time.

Suggestions for Diagnosis

What is meant by diagnosis? By diagnosis is meant the taking of certain symptoms that exist and finding out from them what

the trouble is. Such diagnosis would naturally be followed by proper treatment. Each pupil may have a different trouble, and, therefore, require a different treatment. Thus the importance of the individual is brought to the front. There may be two reasons for looking over a set of papers. First, to simply mark a pupil's paper; second, to gain some knowledge which will serve as a basis for specific corrective work. The latter method of correcting work is diagnosis.

There are three principles which underlie any diagnosis of a pupil's paper.

(a) Trouble with the individual pupils cannot be discovered by studying and drilling the children who have examples right. The principle that success comes through failure must be practised here and the teacher will find her richest material for study in the poorest papers.

(b) After selecting the papers with low scores, it is necessary to find out from the paper, or from studying corrective material how the pupil does his work. The fact that most pupils do their work in the same way makes this possible. Thus it will be necessary to work with the individual and so determine the habits he has formed which lead him to do his work in the wrong way.

(c) The most important step of all is the keeping of a record of each pupil's performances. In no other way can one tell whether the pupil is advancing or not.

Conclusions

3. One of the most important phases brought out as a result of giving standard tests has been the emphasis upon and need of a more careful diagnosis of all work in the elementary school. This is possible only when all phases of the topic under consideration are represented. Tests which do not give all the possibilities may point out an important symptom but cannot be used for purposes of diagnosing the trouble of the pupil.

4. After the results of a test have been studied, drill work, adjusted to the nature of the difficulty, should be given. Pupils who are able to do the mechanical work sufficiently well should not be required to take this drill work required of slower pupils. A method

which requires the entire class to do all the drill work is just as inefficient and unpedagogical as is the method which would allow the slow pupil to drift along and do what he could with only class instruction. Any method of diagnosis emphasizes the individual. Depending on averages and citing of averages ignores and covers up the individual.

1. The mechanical parts of the different subjects must be reduced as nearly as possible to automatism.

2. The various mechanical operations in different subjects are not one general habit but made up of a number of specific habits. For example, *there is no one general habit of adding integers but a number of specific habits involving each possible type in this process.*

MENTALITY TESTS FOR COLLEGE FRESHMEN

W. L. UHL

Northwestern University

The aim of this article is to report the results of a series of tests given in the fall of 1916 to a group of one hundred freshmen at Northwestern University. The purpose of giving the tests was to secure norms for freshmen when tested in large groups, in the hope that such norms might be used in making classifications of students.

Descriptions of the Tests Used

a. The Completion Tests. The Trabue Completion Tests K and M were used. By giving two completion tests, it was believed that the results would show the value of the tests for discovering certain abilities of freshmen. Both of the sets which were used are designed for measuring the abilities of persons of high school age or older. The time allowances for K and M were five minutes and three minutes, respectively; otherwise, the standard directions for giving and scoring the tests were followed.

b. The Hard Opposites Tests. For the opposites test a list of twenty words was used. Two minutes were allowed for this test, the students having written their names upon the backs of the test sheets before turning them.

c. The Information Test. A one-hundred-word list was used for this test. About thirty new words were substituted for an equal number of words of Whipple's list. These substitutions were made because too many words of the original list are new to nearly all freshmen. There was also an attempt to represent, by the words, such fields of knowledge as freshmen may reasonably be expected to have explored. An attempt was made to arrange the words so that every fourth word would be a relatively easy one; *e. g.*, base-hit, Anthony Wayne, X-ray, etc., were separated by words which were believed to be of greater difficulty.

Another change in this test has to do with the responses of the subjects. In its original form, the subject was asked to place one of four letters before each word: D, if he could define it as accurately as would the dictionary; E, if he could explain it clearly enough to give one who was not familiar with it some idea of its meaning, although unable to give an exact definition; F, if the word was only roughly familiar, so that he could not use it intelligently;

and N, if the word was entirely new to him. This method admitted of marked over-or under-estimation, by the subject, of his own ability. In order to have a check upon the alleged knowledge of the words marked, high school subjects taking the original Whipple test were asked to define,* on the reverse side of the blank, the first five words marked D and the first five marked E. The following are some of the definitions given:

"Anthony Wayne—an author.

Architrave—an arch formed by trees.

Base-hit—a three base hit meaning the runner hit the ball far enough so he could get back to home before its return.

Caedmon—is one of the cycles of ancient times.

Catalepsy—a disease or sickness of the mind which often causes fits or entire loss of memory.

Chamfer—a kind of medicine.

Clearing-house—a wholesale house that is selling out.

Euclid—a Latin book of translations; a character in Latin; an ancient personage who was a great hero.

Les Miserables—a piece of music."*

With the exceptions of the responses to the words Catalepsy and Clearing-house, it will be seen that there is a clear case of over-estimation which would have been avoided in marking the tests if the subjects had been required to give even brief definitions as responses. In the hope of securing a better type of responses, the following directions were given to the freshmen.

The purpose of this test is to measure very roughly the breadth of your general information. On the sheet of paper which you have, there is a list of words taken from a large number of fields of knowledge. Your part in this test will be to show that you know from what field of knowledge each word is drawn.

(To illustrate the exact method to be used, the following examples were written on the blackboard).

Bryant—literature (poor but better than no response).
—American literature (adequate).

Snowbound—poem
—American poem or Whittier's poem.
(Comments as to responses were made as under Bryant).

Caesar—history.
—Roman History.

punt—sport.
—foot-ball.

Paris—city.
—French city.

C. O. D.—business term (adequate).

dyne—science.
—physics

*Jones, E. E.: An Unpublished Report of a Survey of the Schools of Cambridge, Ill.

Merely state the field of knowledge unless it is easier for you to define the word. Try to use no more than three words for each word of the test—in many cases, one word will be sufficient. Take the words in order, beginning with the first column. Work as rapidly as you can.

This was the last of a series of tests, all of which had been timed tests: as a result, the freshmen thought they were to have but a limited amount of time for this. After ten minutes, it was seen that practically all were through work—nearly all had stopped; a very few worked longer. No one worked more than twelve minutes.

Results of the Tests

a. The Completion Tests. As the two completion tests were given less than five minutes apart, it was supposed that, with few exceptions, each student would make about the same relative score in one test as in the other. It seemed plausible, however, that there would be less difficulty in getting started upon Test M than upon K, owing to the practice effect. It is possible also that a period of three minutes was too short for Test M. A comparison of the results indicates, however, that much the same degree of success was attained in one test as in the other. The number receiving the maximum score of two on each sentence was practically the same; this tends to the conclusion that the students were given sufficient time to do as well in one test as in the other. This conclusion is supported by other considerations. The number attaining a score of one on different sentences was small; with the exception of the scores for sentence 5, in Test M, there was a close similarity between the results of the two tests in this respect. The scores for *any* written, though futile, attempt upon a sentence are the same in the two tests except in case of sentences 6–8. The explanation of this fact is probably that an insufficient amount of time was given in Test M for making futile attempts to complete these more difficult sentences; some influence may possibly be due to the students' unwillingness, by the time these sentences were reached, to make attempts which were unlikely to be successful. The latter explanation seems untenable, however, because many were willing to make futile attempts upon sentence 6.

In Table I are shown the frequencies of scores for the two tests. The maximum scores for K and M are 14 and 16, respectively. It will be seen that all lacked two of more of attaining the maximum score in M, while three lacked only one of attaining the maximum score for K. Either the necessary shortening of the time for M or

TABLE I

Results of the Completion Tests

Scores		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	or more Total
Frequencies	K	0	1	2	2	4	3	13	10	16	11	22	0	13	3	0	0	100
	M	0	0	2	0	5	3	9	12	18	11	19	9	6	2	4	0	100

Medians K 8.9
M 9.1

TABLE II

Results of the Opposites Test

Scores	0–5	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	Total
Frequencies	1	0	1	6	17	17	26	12	11	7	2	100

Median 31.2

the greater difficulty of the last sentences may account for this difference.* The medians for the two tests indicate that M is the easier test even disregarding the difference in time.

So far the tests seem to have yielded results which are quite similar. When we consider the scores of individual students, however, we see that a very large number shifted their positions greatly. The more striking fact is that practically as many made a low score in the first test as in the second; 42 made a higher score in K than in M; 11 made the same; 47 made more in M.

b. The Opposites Test. The results of the opposites test are shown in Table II. The scheme of scoring suggested by King and Gold† was followed. The essential features of the scheme are as follows: (1) for exact opposites a score of 3 is given; (2) for closely allied but not the best, or for words commonly but loosely used as opposites a score of 2 is given; (3) for words of only remotely opposite signification a score of 1 is given; (4) a score of 0 is given for the following kinds of words: random or coined words, phrases, words of different parts of speech, and synonyms. The responses indicate that sufficient time was given as the last words were tried by nearly all of the students, and practically all worked straight through the list.

---

*The values assigned by Trabue do not support the latter possibility. Pp. 26–7, *loc. cit.*

†Irving King and Hugo Gold, *A Tentative Standardization of Certain "Opposite Tests."* Journal of Educational Psychology, Vol. VII, pp. 459–482, Oct., 1916.

c. The Information Test. The information test proved to be the most difficult to score. In selecting words for the test there was an attempt (1) to eliminate the possibility of chance correct reactions; (2) to find words of such character that a satisfactory reaction consisting of only one word would be possible; and (3) to avoid words having merely local or transient significance. In scoring, the reactions were classified as follows: (a) correct (score of 2); (b) not wrong but too general, vague, or far-fetched (score of 1); (c) still more slight relation to the stimulus-word (score of ½); (d) wrong, or inadmissible on account of being arbitrarily considered as too far-fetched (score of 0). The distribution of scores is shown in Table III.

TABLE III

Distribution of Scores for the Opposites Test

Score	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90	90–100	100–110	110–120	120–130	130 or more	Total
Frequencies	1	2	11	14	11	13	16	14	9	3	5	1	0	100

Median 68.4

In the assignment of values, some cases were treated very arbitrarily, as in the case of "religion" as a reaction to "Zionism;" Zionism is an especially poor word for the test as scored because it admits of a response from all subjects if they are willing to use a very general response. "Religion" as a response to Zionism is, therefore, believed to be much easier than "philosophy" as a response to "Stoicism," although both responses were given equal value. In case of the word "Utopia," a higher value was assigned to the response "ideal land" than to "More's poem," the latter being considered as worthy of a score 1, while "English poem" was given a score of ½. In some cases, as, for example, the last-mentioned response, the less definite and more common response, "English literature," was given a score of 2. This distinction was made because the response, "English literature," fulfills all the requirements of the directions for the test. Some responses were given certain values which result from the fact that their stimulus-words are often found outside their precise fields of knowledge; *e. g.*, "spoils system" may as properly be assigned to American history

as to political science according to the directions for the test; also, *Habeas corpus* is quite as properly an historical as a legal term except that the latter is more definite. If objections be raised in regard to the responses, they should be maintained only after consultation with advanced students from several different departments. It is believed that the judgments of a group of graduate students and instructors will necessitate a considerable number of arbitrary assignments of values, owing to the different relations in which certain words have often been met. It is, of course, obvious that many responses require only the merest smattering of acquaintance with the words of the test. Inasmuch as only proper nouns were capitalized, a penalty was fixed for overlooking capitalization as in the case of Millet which was called "grain."

Discussion of Results

In order to compare the results of the tests with those of university courses, the grades of the freshmen in their first semester courses in required English and mathematics were secured. Also, the three mathematics instructors who had these one-hundred students in mathematics were asked to rank their own students as to general ability. For this ranking, the following simple instructions were given:

> So far as you are acquainted with the following persons, mark down after each name a number indicating the intellectual fitness of each individual. Grade in terms of what you consider to be their real ability or capacity to do difficult intellectual work—not in terms of their customary class performances.
>
> Mark +2 after names of superior individuals;
> Mark +1 after names of very good individuals;
> Mark 0 after names of merely average individuals;
> Mark −1 after names of individuals just below average;
> Mark −2 after names of very poor individuals.

In addition to these checks upon the tests, the university registrar supplied data showing the standing of the freshmen in the high school classes in which they were graduated.

In Table IV are the correlations between the different tests and between the tests and English and mathematics, abilities as judged by instructors, and ranks in high school classes.* The most strik-

*The correlations given in this table were secured by means of the Spearman "footrule for correlation." The values for "R" are here given in terms of values given in Rugg's *Statistical Methods Applied to Education*, page 402.

ing correlation is that between Completion Tests K and M. In case of a relatively homogeneous group of subjects such as university freshmen may be assumed to be, many shiftings of ranks may be expected, but, even so, one would expect a higher correlation in case of similar tests which are considerably different. That is, if the correlation between tests K and M is only .32, it would not be expected that test M and the information test should have a correlation of .42, and then that K and the information test should have a correlation of only .18. The instability of ranks is again

TABLE IV

Correlations between Tests and University Grades

	Compl. K	Compl. M	Oppo-sites	Informa-tion	English	Mathe-matics
Compl. K.....	...	.32	.21	.18	.37	.16
Compl. M....	.32	...	.26	.42	.38	.41
Opposites.....	.21	.26	...	.26	.44	.29
Information...	.18	.42	.26	...	.38	.43
English.......	.37	.38	.44	.38	...	.48
Mathematics..	.16	.41	.29	.43	.48	...

(K and M) and English	.39
(K and M) and Mathematics	.36
Instructors' Judgments and Mathematics	.93
Judgments and High School Ranks	.59
Mathematics and High School Ranks	.55
Mathematics and English	.48
High School Ranks and All Tests	.37
Mathematics and All Tests	.41
English and All Tests	.46
Judgments and All Tests	.36

shown in the case of K and M in their respective correlations of .16 and .41 with mathematics; when the scores for K and M were combined, the correlation with mathematics was .36.

A much higher correlation was expected between the completion tests and English, and, in order to find out whether the low correlation was due to the fact that the tests were as a whole too easy, the scores received by the students upon the last four sentences of K were found. These scores were as follows:

Score	O	19	Score 5	2
	1	4	6	10
	2	18	7	3
	3	14	8	0
	4	30		

This part of the test seems to be too hard as is shown by the bunching of the scores at the lower end. The correlation for this array with English was .29, even lower than that for K in its entirety.

Table V shows how marked were the individual shiftings in the different tests and university courses. The term "quarter" denotes the positions of the *ranks* of students; for example, students whose ranks fell between 1 and 25 were classified as being within the first quarter, between 25 and 50 as the second quarter, etc. This differs from the quartile divisions used in part in the next two tables, because it often happened that the rank of a large number of students was so placed as to throw too many into a given quarter; for example, too many fall into the second quarter for K because a large group received the rank of 25.5. A comparison of various forms of groupings indicated that for our purposes the grouping by quarters is quite as valid and satisfactory as grouping by quartiles. Either form of grouping is arbitrary here, because in several of the tests and in the university courses quartile divisions would not be based upon scores or grades alone, owing to the "crowding" of certain parts of the distribution.*

Table VI shows the degree of accuracy which may be expected in the results of the tests. Division A of this table shows that only a small number of those who received grades of B or C were displaced as much as two quarters; that is, of those who were placed in quartile 3 by Test K, four received a grade of either A or B in English. (It will be remembered that there were only sixteen A's in English). Also, eleven of quartile 1 in K received either C, D, or F in English. These comparisons are made with English because the test is a language test and because the comparisons are still more unfavorable for the test when made with mathematics.

Before comparing the other tests, let us note the relation between mathematics and English, Division D of Table VI. Of those who are in quartile 1 of English, only two are in the third or fourth quarters in mathematics. Now, turning to K and M, we find many more shiftings except for quartile 3. These findings would tend to the conclusion that if one desired to test a student's ability to pursue English successfully, much more precise results would be se-

*There was the following distribution of grades in English: A, 16; B, 27; C, 49; D, 2; F, 6. Quartile I includes nine who received a grade of B.

TABLE V

Relative Positions of Individuals in Different T sts

A. With reference to K.

Quarter in K	Test	Position in different quarters in other tests			
		Quarters in other tests			
	M	1st	2d	3d	4th
1		4	7	2	1
2		10	10	6	7
3		5	10	5	7
4		3	3	7	13
1	Opp.	1	9	2	2
2		9	9	8	7
3		8	7	6	6
4		3	4	7	12
1	Inf.	3	4	3	4
2		10	8	10	5
3		5	5	12	5
4		3	6	7	10
1	Eng.	2	6	6	0
2		6	11	14	2
3		6	3	17	1
4		2	7	12	5
1	Math.	6	4	3	1
2		13	9	5	6
3		10	6	5	6
4		8	10	3	5

B. With reference to English.

Quarter in Eng.	Test	Position in different quarters in other tests			
	M	1st	2d	3d	4th
1		3	8	4	1
2		9	8	6	4
3		9	14	8	18
4		1	1	0	6
1	Opp.	7	4	1	4
2		5	13	5	4
3		7	12	15	15
4		2	1	2	3
1	Inf.	5	5	3	3
2		8	6	6	7
3		10	12	16	11
4		1	0	2	5
1	K	2	6	6	2
2		6	11	3	7
3		6	14	17	12
4		0	2	1	5
1	Math.	7	7	1	1
2		18	8	1	0
3		12	13	12	12
4		0	2	1	5

cured if the student were given a semester's work in mathematics than if he were given either K or M or both. In fact, the shiftings are so great in all the tests that from one-fifth to one-third are displaced as much as would be indicated in terms of university grades by two or three letters; for example, from a grade of A as shown by one test to a grade of C or D as shown by another test. Table VII shows, however, that the method of scoring may have much to do with the shiftings between English and the completion tests.

TABLE VI

This Table Shows the Number of Individuals Who Were Displaced Two or More Quarters by Certain Tests

A. With reference to K.						B. With reference to M.						
Quartile in K	M	Opp.	Inf.	Eng.	Math.	Quartile in M	K	Opp.	Inf.	Eng.	Math.	Judg.
1	3	8	11	11	7	1	9	8	7	11	6	6
2	7	6	5	1	3	2	2	5	7	1	3	2
3	4	7	8	4	9	3	2	4	6	7	10	14
4	5	8	7	9	17	4	8	8	7	3	10	11
Total	19	29	31	25	36		21	25	27	22	29	33

C. With reference to K and M combined.								D. With reference to English.					
Quartile in K–M	K	M	Opp.	Inf.	Eng.	Math.	Judg.	Quartile in Eng.	K	M	Opp.	Inf.	Math.
1	2	0	8	7	10	8	8	1	8	7	7	9	2
2	3	1	4	7	0	2	2	2	14	11	8	7	2
3	0	0	3	4	2	9	9	3	0	4	5	5	9
4	1	1	9	7	8	13	14	4	19	12	9	7	9
Total	6	2	24	25	20	32	33		41	34	29	28	22

In case of the completion tests, grades have been assigned to the respective scores as follows: Scores for the two tests may be from 0 to 30; the scores were, therefore, divided into five equal parts and assigned grades corresponding to their respective positions. This table shows a greater value for the completion tests than is shown by any of the other tables. The correlation (Pearson formula) for mathematics and English is .48, while that for English

TABLE VII

Correlation Charts for English and Mathematics and for English and the Completion Tests

A. English and Mathematics

Grades in English

		A	B	C	D	F	Total
Mathe-	A	3	9	0	0	0	12
matics	B	5	9	11	0	0	25
	C	6	8	17	1	2	34
	D	1	1	10	0	0	12
	F	1	0	11	1	4	17
Totals		16	27	49	2	6	100

B. English and the Completion Tests combined scores of K and M

Grades in English

Scores			A	B	C	D	F	Total
24–29	99	A	0	1	1	0	0	2
18–23	99	B	11	15	17	0	2	45
12–17	99	C	5	9	24	0	0	38
6–11	99	D	0	2	7	2	3	14
0– 5	99	F	0	0	0	0	1	1
Totals			16	27	49	2	6	100

and the completion tests with this distribution is only .27. Even with this low correlation, it will be seen from Table VIII that there are fewer displacements from the grades in English in the case of the completion tests than in the case of mathematics. The results as shown in this table indicate that an instructor might, by

TABLE VIII

Displacements from Grades in English

Grades in English

	A	B	C	D	F	Total
K and M	5	2	1	0	2	10
Math.	8	1	11	0	2	22

a proper method of adjusting his scoring of both his regular class work and the two completion tests, form a fairly accurate opinion of the abilities of his students from such tests alone. Tables V to VIII also indicate very clearly that coefficients of correlation conceal the very things for which an instructor is looking, that is, individual achievements.

In Table IX are given the correlations between some of the tests as indicated by both Spearman and Pearson formulae. While there are considerable differences in case of several pairs of coefficients, the range for the entire group is about the same.

In Table X are shown the ranks of a few individuals in all tests. Division A shows the ranks in other tests of those students who ranked from 1–5 and from 96–100 in K and M combined; as will

TABLE IX

Correlation Coefficients as Shown by Different Formuale.

	Spearman	Pearson
Completion Tests K and M	.32	.38
Test M and Opposites	.26	.42
Test M and English	.38	.34
Test K and English	.37	.24
Opposites and English	.48	.30
Tests K and M and English	.40	.27*
English and Mathematics	.48	.48

*Based upon Table VII.

be seen, even these extremes are not rigidly placed by the tests. Division D is probably the most valuable part of this table. The students who ranked 1 and 97, respectively, are good examples of those who are likely to be misplaced by brief tests. In individual tests given to them later, No. 1 was alert and rapid except in the Healy Puzzle-box test where both were very slow. No. 97 showed less interest in the tests and worked them more slowly. In regard to the university grades, No. 1 resented the opinion that there might be a close relation between scores made in tests and class grades; she said that she did not like mathematics nor her instructor. She appears to be the type that will respond readily and rapidly to a social situation. On the other hand, No. 97 is far more phlegmatic; also, her instructor ranks her very high in general ability. One of these subjects thus appears to be able in short tests to surpass her long-continued efforts, while the other can do little unless given ample time for reflection.

TABLE X

Ranks in All Tests of Those Receiving Very High or Very Low Scores

DIVISION A

Ranks in K and M	Ranks in other tests				Grades	
	K	M	Opp.	Inf.	Eng.	Math.
1	2	2.5	39	5.5	B	A
2	9	2.5	12.5	12	C	D
3	42	2.5	88	49.5	B	A
4	9	18	43	48	A	F
5	2	32	35.5	16.5	B	B
96	93.5	92.5	2	30	C	D
97	98.5	87	49	54.5	B	A
98	100	76.5	75.5	7	C	C
99	93.5	96	100	98	F	F
100	98.5	96	16.5	66	F	C

DIVISION B

Ranks in Opposites	Ranks in other tests				Grades	
	K	M	K & M	Inf.	Eng.	Math.
1	42	10	17.5	47	C	B
2	93.5	92.5	96	30	C	D
3	25.5	18	17.5	93	B	A
4	55.5	87	77.5	60	C	F
5	55.5	76.5	72	8	B	A
96	25.5	47	32.5	70	C	B
97	25.5	76.5	54	58	C	C
98	81	61.5	71.5	69	C	C
99	25.5	47	32.5	76.5	C	D
100	93.5	96	98	98	F	F

DIVISION C

Ranks in Inf.	Ranks in other tests				Grades	
	K	M	K & M	Opp.	Eng.	Math.
1	55.5	32	42.5	69.5	C	F
2	55.5	87	77.5	7	A	A
3	42	18	24.5	16.5	B	B
4	69	5.5	24.5	69.5	F	C
5	2	2.5	1	39	B	A
96	42	87	72	69.5	C	C
97	25.5	32	24.5	20	F	F
98	93.5	96	98	100	F	F
99	42	32	32.5	20	B	A
100	42	98	87	94	F	F

DIVISION D

Ranks in All Tests	Ranks in each test					Grades	
	K	M	K & M	Opp.	Inf.	Eng.	Math.
1	9	2.5	2	12.5	12	C	E
2	2	2.5	1	39	5.5	B	A
3	25.5	10	11	20	14.5	B	C
4	25.5	18	17.5	25.5	5.5	B	B
5	42	18	24.5	16.5	25	B	B
96	93.5	87	92	75.5	83	C	C
97	81	92.5	87	78	91.5	B	A
98	81	87	85	86.5	88.5	C	F
99	81	96	92	84.5	99	F	F
100	93.5	96	98	100	98	F	F

Summary and Conclusions

1. As is true of tests in general, these tests fail to measure both accurately and adequately. While a test may call for the exercise of a general ability, yet it can call forth only a slight or fragmentary type of reaction; as a consequence, further testing by means of a similar test may give quite different results. This fact is illustrated in case of the completion tests. It seems likely that tests similar to the information test might give very different results.

2. If, however, standardized tests can be made adequate, they will offer a superior means for testing general abilities. An instructor can devise "quizzes" which test both general and special abilities; these are open to the objection that they cannot well be standardized as to either the method of giving or the amount of demand made upon general as distinguished from special preparation. A standardized test should reveal either special strength or weakness; a form of weakness may be that of being unable to do well in a brief intense test, or inability to use language forms accurately and automatically. The completion tests if given upon several different days in several different forms should reveal such weaknesses. By means of such a form of comparison as that shown in Table VII it was seen that there was as much likelihood of arriving at approximately correct judgments of the students tested by noting their achievements in the completion tests as in mathematics.

3. In order to make the results most reliable, individual testing may be preferable, but distributed periods as just suggested seem, from the results of these tests to be satisfactory so far as group work is concerned.

4. The homogeneity of the group upon which this report is based probably has much to do with the failure to secure high correlations. In such a group there is more competition for the same places or ranks than is true in case of a group in which some members are very far below and others far above the median. While there is marked difference between the ability of a student who attains high marks and one who attains low marks, there is a far greater difference between the extremes in the sixth grade of elementary school, that is, before highly selective processes have been long at work.

5. Of the tests given, the information test seems the least practicable. Such a test might be very valuable in giving an index of ability in any given field of knowledge; as a general test, however, its results are vague, hard to evaluate, and show no higher correla-

tion than the other tests. It has, of course, the advantage of indicating something that is not shown by any of the other tests. More careful selection of words from different fields and a standardized method of scoring are necessary if the test is to be of greatest value.

6. In attempting to devise tests and to secure standards which will show a high correlation with school work, one faces the problem of testing in a brief interval of time those special or general capacities that determine one's ability to carry on successfully a very complex type of activity for an extended period of time and under varying conditions. The problem seems to be too complicated to be explained merely upon the basis that the tests measure more highly specialized capacities or different capacities than those found in a university course such as English. For example, the concomitance of abilities to succeed in English and mathematics is more frequent than that of abilities to do well in tests involving the use of English forms and the university course in English. While mere concomitance of abilities may show no necessary relation between them, yet it is, on the other hand, possible that problem-solving of one kind or another enters into the two university subjects sufficiently to make their demands more nearly the same than the demands of English and the tests. The social situations involved in such courses are certainly much more nearly the same than those in tests and regular work.

7. Again, the low correlations may be due in part to the relative simplicity of tests as compared with complex courses. Besides the social situations just mentioned there is in one a long-continued demand upon one's will to preserve in case of a university course, while a test is a novel form of activity which, though often seeming quite impossible, is of very short duration. There must be ability to set one's self promptly at the test in a way which the semester's work may not require; this was shown in case of student Number 97 in Division D of Table X. The effect of seeing one's neighbors finishing a test probably has the effect of interfering greatly with the forms of automatic responses and ability to discriminate as demanded by the test. One of the great difficulties in working through a test is often the fact that one becomes obsessed with an idea in the form of a word or phrase; he then tries to force a solution. In this period of comparative destitution, any form of distraction may cause failure.

8. At the outset of this paper, it was said that one of the aims was to try a method of testing; the results secured seem to warrant the use of the method in case of tests in which a standard amount of time is to be given to all. Monitors are, of course, needed to prevent efforts to work overtime. As to the practicability of the tests, the completion tests, and the opposites seem to be better adapted for such use than the information test on account of reasons stated under "5." It is believed that the instructor would secure valuable information concerning his students' abilities by studying the results of this kind of tests. The medians for the tests are shown in Tables I–III and may be considered as norms. The fact that either weak or strong students may, by the amount of effort which they put forth, secure high marks in university courses, makes impossible an absolute judgment as to one's ability; for this reason, the results of tests may be of value in planning needed assistance to certain students thus aiding in checking the appalling "mortality" of college freshmen.

SOME DATA ON THE BINET TEST OF NAMING WORDS*

LEWIS M. TERMAN

Leland Stanford, Jr. University

Some of the questions which suggest themselves regarding the 60 word test are the following:

1. To what extent does the test correlate with other measures of intelligence?

2. What number of minutes alloted to the test will yield the best measure of intelligence?

3. Does the slope of the curve for words named in the six successive half minutes differ for children of different degrees of brightness?

4. To what extent is the ability to name words in a free association test dependent on size of vocabulary?

5. Do older and younger children of the same mental age differ in the number of words named?

6. What sex difference does the test reveal?

7. What is the significance of such features of the response as stereotypy, improvisation of words, the size of the association groups, the nature of the associations involved, the degree of abstractness of words used, etc.

Questions relating to the psychological analysis of the response are especially interesting and important, but the writer is not prepared to deal with them here. The present paper will be limited to the presentation of certain quantitative data bearing chiefly on the value of the test as a measure of mental ability.

Correlation with Mental Age

Table I shows for a miscellaneous group of 480 children the correlation between mental age and number of words named in three minutes.

The correlation of .535 is considerable, but not as high as might have been expected. Had there been a larger proportion of our subjects of extremely high and also of extremely low mental age, the correlation would have been higher. The large majority of subjects entering into Table I are between 8 and 13 in mental age.

*The writer is indebted to Miss Bess Henry.

TABLE I

Total Words in 3 Minutes

Mental Age	10	15	20	25	30	35	40	45	50	55	60	65	70	75	80	85	90	95	100	105	110	115	120	125	130	135	Total
17^6																		1								1	2
17																											0
16^6																				1	1	2					4
16												1															1
15^6												1		1								1					3
15													1		1			3									5
14^6													1		1	1		2									5
14													2	1	1	1											5
13^6											1		3	3	2		2	1				1					13
13											4	4	2	2	1	1	2	3	2	2					1		24
12^6						1			2		2	3	4	4	2	2		1	2						1	1	25
12								1	3	2	5	7	4	2	2	3	1		1	1							32
11^6						3	1		1	1	3		1	1	6	3	1			1							22
11							2	2	1	4	6	6	9	5	2	4	1	3	1			1			1		48
10^6						1		1	1	10	7	6	6	1	1	2		3	1		1		1				42
10			1		1	1	2	3	10	8	4	3	6	3	4	2	4	1									53
9^6	1			1		1	1	3	3	5	5	11	1	1	2	2		2		1	1		1				42
9				1	2	4	2	6	6	7	6	3	4	3	3	1		1		1							50
8^6		1	1	1	4	3	3	2	8	6	4	4	1	2	3			1									44
8			2	1	1	3	3	5	3		4	5	2		3		1										33
7^6		1			2	2	3			1	1																10
7			1		1		3	2	2	4																	13
6^6		1				1			1																		3
6					1																						1
Total	1	3	5	4	12	20	20	25	41	48	52	54	47	29	34	22	12	22	7	7	3	5	2	0	3	2	480

r=.535

Note—Columns include scores 0–4, 5–9, 10–14, etc. Rows include mental age ranges of one-half year: 6–6 equals 6-6 to 6-12, etc.

Since the group is of limited heterogeneity, a better index of the value of the test can be found by computing the probable error of a mental age based on this one test. Treating separately the mental age distributions of those naming 30 to 39 words, 40 to 49, 50 to 59, etc. we get the following:

TABLE II

Score	Median mental age Stanford-Binet	P. E.
30–39	8 years 10 months	8.25 months.
40–49	9 years 0 months	10.7 months
50–59	9 years 11.2 months	9.9 months
60–69	10 years 7.4 months	15.2 months
70–79	11 years 4 months	15.2 months
80–89	11 years 3 months	14.4 months
90–99	11 years 6 months	10.5 months
	Average P. E.	12.01 months

The average P. E. of a mental age derived from the 60 word test is almost exactly one year, or at least this is the case when the number of words named is between 30 and 100. Below 30 and above 100 our subjects are too few to give data of value for this purpose. With a P. E. of 12 months, the test may be expected to give a mental

age correct within 18 months in 69 per cent. of cases, and within 24 months in 82 per cent. of cases. The 60 word test is a somewhat less accurate measure than the vocabulary test, which in the case of children, gave a P. E. of 9½ months in terms of mental age. However, it is not as inferior as would appear from a comparison of the coefficients of correlation with mental age (.91 and .535).

THE RELIABILITY OF A ONE-MINUTE TEST

One would naturally expect much greater accuracy for a three-minute than for a one-minute test. To test this assumption the score for the first minute of the test has been separately correlated with mental age. The results are shown in Table III.

TABLE III

Total Words in One Minute

Mental Age	0	5	10	15	20	25	30	35	40	45	50	55	60	65	70	Total
17⁶	.	.	..	..	..	..	...	..	..	..	1	..	..	..	..	1
17	.	.	..	..	..	..	...	..	..	..	..	..	..	..	..	0
16⁶	.	.	..	..	..	..	...	..	..	2	..	1	..	..	..	3
16	.	.	..	..	..	1	...	..	..	1	..	..	..	..	..	2
15⁶	.	.	..	..	..	..	1	..	..	..	..	..	..	..	..	1
15	.	.	..	..	..	..	...	2	3	..	..	..	..	..	..	5
14⁶	.	.	..	..	..	..	3	1	1	..	..	..	..	..	..	5
14	.	.	..	..	..	..	1	2	2	..	..	..	..	..	..	5
13⁶	.	.	..	..	..	..	3	6	3	..	..	1	..	..	..	13
13	.	.	..	..	2	2	6	6	4	2	..	..	1	..	..	23
12⁶	.	.	..	..	..	3	3	8	3	3	1	1	..	..	..	22
12	.	.	..	1	3	8	8	10	1	2	1	..	..	..	..	34
11⁶	.	.	1	1	2	2	5	8	..	1	..	..	..	..	..	20
11	.	.	1	3	7	8	9	8	4	3	3	..	..	..	..	46
10⁶	.	.	..	1	2	10	9	5	5	3	..	..	..	..	..	35
10	.	.	2	4	8	14	13	11	3	..	..	..	..	..	..	55
9⁶	.	1	..	3	10	7	11	6	3	1	..	..	..	..	..	42
9	.	.	1	8	5	11	12	7	1	..	..	..	..	..	..	45
8⁶	1	.	5	4	3	9	10	4	1	..	..	..	..	..	..	37
8	.	.	3	4	7	6	7	3	..	..	..	..	..	..	..	30
7⁶	.	.	3	..	6	..	1	..	..	..	..	..	..	..	..	10
7	.	.	1	4	4	3	...	..	..	..	..	..	..	..	..	12
6⁶	.	.	2	1	..	..	...	..	..	..	..	..	..	..	..	3
6	.	.	..	..	.	..	...	..	..	..	..	..	..	..	..	0
5⁶	.	.	..	1	..	..	...	..	..	..	..	..	..	..	..	1
Total	1	1	19	35	59	84	102	87	34	18	6	3	1	..	..	450

r = .52

Note—Columns include scores 0–4, 5–9, 10–14 etc. Rows include mental age ranges of one-half year: 6–6 equals 6–6 to 6–12, etc.

The writer was surprised to find that one minute constitutes as good a measure in this test as three minutes. The correlation with mental age is almost exactly the same in the two cases, (.535 and .52). The P. E. of mental age based on the one-minute test is approximately 13 months; or only a month higher than for the three minute test. The saving of two minutes would be material and it is proposed that the test be shortened to one minute and that 28 words be required for a pass at year X instead of 60 words in three minutes.

The correlation between the first minute and the last minute was computed and found to be .58.

The Influence of Foreign Language in the Home

The correlation with mental age was computed separately for 102 Portuguese and Italian school children. Nearly all of these were from homes in which a foreign language is spoken, although a majority of the children had learned English before entering school. The correlation was .43, which is slightly lower than for American children. The medians for the different mental ages did not, however, differ materially in the two groups. (See Table IV.) The conclusion would be that the test is not seriously enough vitiated by the language factor to justify its omission in testing children of foreign parentage, provided the children have spoken English for three or four years.

Decrease of Words Named in Successive Minutes

Theoretically, the percentage of decrease in number of words named in the successive minutes might be significant. Observation had led the writer to believe that average and superior children would show a smaller percentage of drop from the first minute to the second and from the second to the third, than do dull and feeble-minded subjects. Accordingly the percentage of drop was calculated for 310 children and the results were then correlated with intelligence quotient. This was done for the 180 boys and 130 girls separately.

The results show how erroneous an impression is likely to be when based on nothing but casual observation. The percentage of drop from the first minute correlated with intelligence quotient to the extent of only .227 with boys and .19 with girls. The percentage of drop from the second minute to the third minute correlated with intelligence quotient to the extent of .185 with boys and .135 with girls.

The children were then divided into three groups according to intelligence quotient: (1) those below 80; (2) those between 80 and 110; and (3) those above 110. The median percentage of drop for the three groups was as follows:

	First minute to second minute		Second minute to third minute	
	Boys	Girls	Boys	Girls
IQ below 80	40%	46%	51.2%	54%
IQ 80–110	31.5%	35%	46.7%	41.3%
IQ above 110	33.7%	31.7%	37.5%	43%

Although the coefficient of correlation is low, there is a consistent tendency for bright children to maintain their initial speed considerably better than do the dull. There are so many exceptions, however, that the amount of drop cannot be regarded as a dependable symptom in clinical work.

Incidentally it is interesting to note that the third minute is about as inferior to the second as the second is to the first; also that in percentage of drop there are no significant sex differences.

Increase by Mental Age in the Ability Tested

Table IV shows for each mental age the median number of words named in three minutes by the sexes taken together, by the boys and girls taken separately, and by the Portuguese-Italian group; also the medians for the number named in one minute (American group).

TABLE IV

	Three Minutes				One Minute
Mental age	Sexes together	Boys	Girls	Port-Italian	American group sexes together
13-6 to 14-5	78.8	77	83.3	81	38.1
12-6 to 13-5	77.5	72.5	79.5	79.	37.1
11-6 to 12-5	69.5	64.1	72	77	33.5
10-6 to 11-5	69.2	68.3	69.4	70.5	32.8
9-6 to 10-5	64.4	60.8	67.5	62.5	30
8-6 to 9-5	56.5	61.1	57.5	56.7	28.5
7-6 to 8-5	49	43.7	45	53.5	23.8
6-6 to 7-5	47.5	47.5	49	50	19

FIGURE I. Mental age increase in number of words named in three minutes. (American children, both sexes.)

Dependence Upon Vocabulary

For 360 children the correlation between vocabulary and number of words named in three minutes was .487. The two abilities are evidently fairly closely related. This is probably due to their mutual dependence upon intelligence. It is not known whether children of the same mental age but of widely different vocabulary would differ greatly in the 60 word test.

Criticism of the Test

In the first place the writer would urge the desirability of working out a satisfactory method of giving credit for quality as well as for quantity. This is important but will be a difficult task.

Performance in the test is unquestionably influenced by the conditions under which it is given and by the surroundings. In view of the impossibility of providing a standard room, containing always a given equipment, it has been suggested that the subject be required to close the eyes. It would be interesting to see whether this procedure would increase the correlation with mental age.

The performance is also influenced by the amount of emphasis placed by the examiner upon speed. In *The Measurement of Intelligence* the writer has indicated the degree of emphasis which he regards permissible.

Finally, the test is by no means easy to score. When words are named at the rate of 20 to 40 per minute it is hard to keep count, especially since the examiner must also watch the time, note repetitions, and pay some attention to the order of associations and quality of performance. The test is harder work for the examiner, if he is conscientious, than for the subject. As it is ordinarily not possible to record the words unless one can write in shorthand, the record should be made by pencil marks. Even then a class of students recording a demonstration test will often vary in the count as much as three or four words if the subject is rapid. The disagreements are largely caused by the neglect to eliminate repetitions and to confusion in recording compound words. It is recommended that words like "door knob," "sulphuric acid" etc., should be counted as one.

Summary

1. The 60 word test is, for a single test, a fairly good measure of intelligence. In 50 per cent of cases it yields a mental age within a year of that secured by the Stanford-Binet Scale.

2. The time can be reduced to one minute without appreciable loss of accuracy, as a one-minute test yields a mental age whose P. E. is only 13 months.

3. In the case of children who have attended school for three or four years the test is not seriously vitiated by the use of a foreign language in the home.

4. The percentage of drop in the number of words named in successive minutes is greater for dull than for bright subjects, but the correlation is low.

5. The percentage of drop from the first minute to the second is about the same as from the second minute to the third.

6. The ability tested increases fairly regularly in successive mental ages.

7. Special care is necessary to secure accuracy and uniformity of scoring.

A PRELIMINARY TEST IN CHEMISTRY

HANOR A. WEBB

Department of Chemistry, George Peabody College for Teachers

Many types of educational problems are being studied from quantitative data. The value of standard tests in elementary school subjects has encouraged similar experiments in high school classes, and one by one "tests" for the various secondary subjects are published.

A clear distinction has been recently expressed between tests of skill in performance, and tests of range of information.† Performance, both in speed and accuracy, is most conveniently measured by means of a series of problems to be solved within definite time limits—range of information is determined with a somewhat similar series of queries concerning facts which, in general, the student either knows, or does not know. In connection with the subject of chemistry, the distinction between problem and informational tests may be more clearly perceived by comparing the preliminary questions prepared respectively by Professors Jones* and Bell.†

A "standard test" is usually designed to compare the results of previous instruction with some properly derived standard,—measurements more relative than absolute. But little has been done towards the determination of the amount of information possessed by the children *before* they start in upon the period of instruction. These so-called "apperceptive masses," upon which the new concepts must necessarily be built, are undoubtedly of infinitely greater qualitative variety, and wider quantitative range, than even the most scattered results of a term's instruction. The question "What do they already know about the subject?" is in every teacher's mind as he faces a beginning class.

This "tare" in weighing the results of instruction must receive attention if the achievements are to be properly interpreted. But on the qualitative side, what the children already know is the teacher's accessory *before* the fact,—flexibility in method or adaptation of subject matter depends on the discovery of the foundations of information already prepared for the new construction.

---

†J. CARLETON BELL. *A Test in First Year Chemistry,* Journal of Educational Psychology, Vol. 9: 1918, 199–209.

*FRANKLIN T. JONES. *Union Tests in Physics* (Chemistry in preparation), Cleveland, O.

One of the first conceptions necessary in understanding elementary chemistry is the essential difference between *elements, mixtures,* and *compounds.* Their specific properties are illustrated in the first experiments,—their definitions are among the first learned. Texts usually consider these three topics of equal difficulty, and equal time is no doubt given to each in most classrooms. But preconceptions and confusions regarding these substances already exist in the minds of students. Which class of matter is most easy to understand? Which most difficult? Do the experiments of the manual and text clearly illustrate their differences?

For some years the writer has been handing a test sheet on elements, mixtures, and compounds to his normal school and college classes in beginning chemistry. At the head of the sheet, definitions for each class of matter were called for,—then an alphabetical list of fifty well-known substances was given,—nine elements, thirty-two mixtures, and nine compounds required classification. The student filled out the paper without assistance *after* having performed the experiments on this topic in the manual, and studied the text. The sheets were handed in *before* class discussion, it being the idea of the instructor to glance through them, note the prevailing misconceptions, and adjust class demonstrations and discussions accordingly

Out of the accumulation of some years, two hundred fifty papers have been selected, in which the definitions of the three classes of matter are correctly given, and in which all of the fifty substances are classified. The complete data of this test is displayed in Table I

THE RESULTS

I. Majority judgments. Table II.

Of 9 elements,	8 correct,	0 incorrect.
Of 32 mixtures,	9 correct,	10 classified as compounds.
Of 9 coumpounds,	8 correct,	0 incorrect.

Majority judgment is correct as to elements. It also judges compounds truly when they are compounds, but is greatly confused by most mixtures, interpreting them as compounds.

II. Plurality judgments. Table II.

Of elements,	1 correct,	0 incorrect
Of mixtures,	3 correct,	10 incorrect.
Of compounds,	1 correct,	0 incorrect.

TABLE I

Rank Order of Substances, as Judged

	Elements		Mixtures		Compounds	
	Iron	201	Lemonade	207	Water	188
	Phosphorous	195	Face Powder	184	Bone	156
	Zinc	190	Chewing Gum	167	Sugar	150
	Sulfur	180	Gunpowder	158	Soda	149
	Diamond	180	Ink	150	Glass	148
	Copper	180	Soil	148	Salt	145
	Lead	172	Soap	129	Hair	144
	Nickel	156	Milk	129	Alcohol	144
	Soot	101	Flour	125	Epsom Salts	143
	Steel	99	Gelatin	122	Lard	142
	Wool	92	Air	116	Coal	141
	Cotton	84	Dyes	113	Eggs	141
Q_3						
	Salt	79	Paper	112	Meat	139
	Granite	77	Butter	110	Rubber	138
	Silk	72	Smoke	100	Wood	138
	Hair	72	Tea	96	Camphor	136
	Brass	69	Camphor	94	Corn	130
	Soda	61	Silk	89	Dyes	130
	Coal	61	Eggs	88	Cotton	123
	Rubber	59	Corn	87	Brass	122
	Chalk	57	Coffee	85	Chalk	120
	Alcohol	56	Granite	82	Coffee	118
	Tea	53	Meat	76	Paper	117
	Wood	51	Lard	74	Soap	117
	Sugar	49	Chalk	73	Steel	115
M						
	Coffee	47	Epsom Salts	72	Butter	115
	Smoke	46	Bone	65	Wool	112
	Glass	43	Wood	61	Gelatin	108
	Meat	35	Glass	59	Air	106
	Epsom Salts	35	Brass	59	Smoke	104
	Lard	34	Soot	57	Milk	102
	Flour	33	Rubber	53	Tea	101
	Corn	33	Sugar	51	Soot	92
	Bone	29	Alcohol	50	Flour	92
	Air	28	Coal	48	Granite	91
	Butter	25	Wool	46	Ink	90
	Paper	21	Cotton	43	Silk	89
	Eggs	21	Water	42	Soil	87
Q_1						
	Water	20	Soda	40	Gunpowder	86
	Gelatin	20	Steel	36	Chewing Gum	78
	Camphor	20	Hair	34	Diamond	65
	Milk	19	Nickel	29	Nickel	65
	Soil	15	Salt	26	Face Powder	63
	Ink	10	Lead	20	Lead	58
	Dyes	7	Zinc	16	Copper	56
	Gunpowder	6	Sulfur	15	Sulfur	55
	Chewing Gum	5	Copper	14	Phosphorus	45
	Soap	4	Phosphorous	10	Zinc	44
	Lemonade	3	Iron	9	Iron	40
	Face Powder	3	Diamond	5	Lemonade	40

Since a substance might be judged as belonging to any of three classes, some did not receive a majority in any class.

The combined majority and plurality judgment places all elements correctly, all compounds correctly, 12 out of 32 mixtures correctly (37.5%), and errs only in classifying 20 mixtures (62.5%) as compounds.

TABLE II

Substances Grouped by Majority and Plurality Judgments

Elements		Mixtures		Compounds	
Iron	201	Lemonade	207	Water	188
Phosphorous	195	Face Powder	184	xBone	156
Zinc	190	Chewing Gum	167	Sugar	150
Copper	180	Gunpowder	158	Soda	149
Diamond	180	Ink	150	xGlass	148
Sulfur	180	Soil	148	Salt	145
Lead	172	Milk	129	xHair	144
Nickel	156	Soap	129	Alcohol	144
		Flour	125	Epsom Salts	143
———	———			xLard	142
Soot	101	———	———	xCoal	141
		Gelatin	122	xEggs	141
		Air	116	xMeat	139
		Silk	89	xRubber	138
				xWood	138
				Camphor	136
				xCorn	130
				Dyes	130
				———	———
				xCotton	123
				xBrass	122
				Chalk	120
				xCoffee	118
				xPaper	117
				xButter	115
				xSteel	115
				xWool	112
				xSmoke	104
				xTea	101
				xGranite	91

The horizontal line shows division between majority and plurality judgments. x indicates incorrect classification.

In discussion, students always question the classification of a few of these substances, where the characteristics are not well defined, or not uniform. For example, soap, though chemically a compound, is commercially a mixture, and is more properly classified as such with beginning students.

TABLE III

Rank Order of Substances as Correctly Judged

M	Lemonade	207	p M	Gelatin	122
E	Iron	201	p C	Chalk	120
E	Phosphorous	195	p M	Air	116
E	Zinc	190	x M	Paper	112
C	Water	188	x M	Butter	110
M	Face Powder	184	p E	Soot	101
E	Copper	180	x M	Smoke	100
E	Diamond	180	x M	Tea	96
E	Sulfur	180	p M	Silk	89
E	Lead	172	x M	Eggs	88
M	Chewing Gum	167	x M	Corn	87
M	Gunpowder	158	x M	Coffee	85
E	Nickel	156	x M	Granite	82
M	Ink	150	x M	Meat	76
C	Sugar	150	x M	Lard	74
C	Soda	149	x M	Bone	65
M	Soil	148	x M	Wood	61
C	Salt	145	x M	Glass	59
C	Alcohol	144	x M	Brass	59
C	Epsom Salts	143	x M	Rubber	53
C	Camphor	136	x M	Coal	48
C	Dyes	130	x M	Wool	46
M	Milk	129	x M	Cotton	43
M	Soap	129	x M	Steel	36
M	Flour	125	x M	Hair	34

The letters E, M, and C, stand for the correct classification as element, mixture, or compound.

All substances in the first column are correct by majority judgment.

In second column, p indicates correct by plurality judgment; x indicates incorrect by plurality judgment.

III. Correct judgments. Table III.

This table represents the "order of familiarity" with the composition of these fifty substances. It is again to be noted that the elements in the list rank high in familiarity. If unfamiliar elements had been chosen, their rank would have been low, but such a selection would have violated the spirit of this test,—all of the fifty substances are familiar objects *per se*, and the variation is purely in the knowledge of their general composition.

The majority of mixtures received few correct judgments.

IV. All judgments. Table I.

Table I. gives three ranked distributions which admit of much statistical treatment. Certain values are perhaps significant in this study of the misconceptions in the minds of students regarding the three classes of matter.

(*a*) *The median.*

Median number of judgments for each classification: elements, 48; mixtures, 72.5; compounds, 115.

Students were more decided in their judgments concerning elements, for half of the substances received 48 or fewer votes. The judgments are most scattered when compounds are considered.

(*b*) *The average.*

Average number of judgments for each substance classified: as element, 64.1; as mixture, 77.5; as compound, 108.3.

The average is a better central tendency to consider in this problem than the median, for the extreme cases are here significant. The students have shown a decided tendency to classify substances as compounds.

(*c*) *The quartile ranges.*

Range of	for 9 elements	for 32 mixtures,	for 9 compounds.
Q_4	1830	1784	1791
Q_3	816	1166	1643
Q_2	430	706	1289
Q_1	132	254	695
Total judgments,	3208	3874	5418
Correct judgments possible.	2250	8000	2250

The significant values here are not so much the quarter points, as the number of measures (judgments) in each quartile range. The concentration of judgments in the upper ranges of the element group is in marked contrast to the scattering throughout all ranges of the compound group. The element column denotes a confidence and definiteness of opinion; the compound column indicates indecision,—the feeling, "If I don't know what it is, I'll call it a compound." And from this, the mixtures have suffered most. Not only have plurality judgments shown the tendency to classify mixtures as compounds, but this tendency is here shown to be of

major proportions. Of the 8000 correct judgments which the mixtures might have received, 3618 went *unpaired* into the compound column.

(*d*) *The extreme measures.*

No. of judgments,	elements,	mixtures,	compounds.
Highest,	201	207	188
Lowest,	3	5	40

The principal confusion is again shown to lie in the compound class. That only 188 (75.2%) correctly understood the nature of water, and that 40 (16%) erroneously classified iron and lemonade as compounds, shows that the prevalent conception is most hazy.

V. Measures of relationship.

(*a*) *Nature and per cent. of errors.* Data from Table I. shows the following relationships of true and erroneous judgments.

Substance	Classified as		
	element,	mixture,	compound
Elements,	69.1%	7.8%	23.1%
Mixtures,	15.9%	39.2%	44.9%
Compounds,	17.1%	24.9%	58.0%

Thus, elements are the best, and mixtures the least understood. The largest error is in classifying mixtures as compounds,—the smallest error, classifying elements as mixtures. The intermediate errors are of moderate size.

(*b*) *Coefficients of correlation.*

As the judgments of two hundred fifty students has given each of fifty substances a rank in three separate classifications, the correspondence of these ranks, (taking into full account the actua number of judgments for each substance), is best expressed by the coefficient of correlation. The best correlation in these cases would be a decidedly minus value, for if many substances were correctly judged, they should rank high in their own classification column, and low under each of the other heads. A strongly negative correlation shows that the properties which distinguish two classes are clearly understood by students, while a smaller negative numerical value shows that there has been much haphazard swapping of judgments. The coefficients of correlation*

between elements and mixtures is −.790
between elements and compounds, −.555
between mixtures and compounds, −.063

*Calculated by "Pearson's formula." The rather complex method is clearly outlined in *Statistical Methods Applied to Education*, by H. O. Rugg, pp. 251–270.

Conclusions

In learning to distinguish between elements, mixtures, and compounds, there is one outstanding, predominent error of judgment made by beginners in chemistry,—the classification of mixtures as compounds. An instructor must earnestly supplement the experiments of manual and text at this point. Other misjudgments are no greater in quantity than would be expected of a normal class.

At least two important causes of error are apparent:

(a) The student has relied on the physical appearance of *homogeneity,*—glass, brass, rubber, steel, etc., seem to be uniform in structure, like compounds. The typical experiments, with iron filings and sulfur, sugar and sand, etc., have been inadequate to clear up this point.

(b) Most students have little conception of the cell structure of plant and animal tissues. Meat, bone, wood, hair, etc., being cellular, are necessarily mixtures.

This study emphasizes the great need for a more scientific attitude in relation to the common things of daily life, and is possibly an argument for some type of science instruction in the grades. That less than half of two hundred fifty nearly grown young men and women knew that air was a mixture makes us wonder how simple a scientific fact must be to become a matter of common knowledge under present methods of instruction. Teaching the three classes of matter is but one of the many problems of beginning chemistry where it is well to know the extent of previous knowledge, right and wrong, and if the standard test, arranged for a given subject, can portray to us the content of a student mind "after taking," let efforts be made to complete the contrast, and show us what was there "before."

COMMUNICATIONS AND DISCUSSIONS

THE EFFECT OF THE SUMMER VACATION ON ABILITY IN THE FUNDAMENTALS OF ARITHMETIC

This investigation was undertaken primarily for the purpose of ascertaining the influence that the summer vacation has on arimetical ability, with a view to locating weak points needing emphasis on return to school. Inasmuch as the school has a cosmopolitan population including children of varying economic and social environment, three distinct types of vacations were recognized; the vacation spent in play, the vacation spent in work and the vacation spent in study. The comparative influences of these three types of vacations on the ability in fundamentals were studied.

The measuring instrument used was the Courtis Test in Fundamentals, Series B. Form 1 was used in June 1918 and Form 2 was used in September 1918. All four tests, that is, addition, subtraction, multiplication and division were given.

Terms Used

"Play Group"—Children whose vacation was passed doing whatever the child desired, but not in working or studying to make up a grade.

"Work Group"—Those children whose vacation was passed in employment. Unless the child worked at least four weeks the vacation was not classed as "work." A child employed to help parent in store, etc., for only two or three hours a day was not considered in this group.

"Study Group"—Those children whose vacation was spent in studying at a summer continuation school to make up a grade.

Scope of Test

The test was given to all 5A, 5B, 6A, 6B, 7A, 7B children in Public School 9, The Bronx. 861 children of whom about 1–3 were girls were tested in June. Owing to transfers the number of these same children in September was reduced to 747. The "work" and "study" groups remained constant, only those children being included who took the test both in June and September. In tabulating the results no distinction was made between boys and girls, because no attempt was made to study the relative ability of the sexes.

Conduct of the Test

The children were taken into an assembly room seating eight classes at a time. The tests were all conducted by the writer personally. The directions that appear at the top of each test sheet were read according to instructions and the prescribed time given for work. All four tests were given at the same session, intervals

of a minute or two being allowed between tests for stretching and relaxation. The entire test consumed about forty minutes for each group of classes.

The weather on all occasions in June and September was about the same; fairly warm but comfortable. In September the humidity was somewhat greater than in June making it slightly oppressive to some of the children. The difference, however, was not sufficient to make it a consideration.

The tests were corrected by a 6B class trained for the purpose and under the supervision of the writer and the class teacher.

Each class teacher tabulated her class results according to instructions given by the writer. A conference was held for the purpose. Those tabulations were checked by the writer and then summarized by him.

Although the primary interest was in the comparative abilities in June and September of the "work," play" and "study" groups the results were also summarized by grade groups so that a grade standard could be derived for purposes of comparison. In the present study the grade groups were combined into year groups as that was deemed more satisfactory for the purpose in view.

Unfortunately the "work" and "study" groups in each year did not prove large enough to warrant comparisons being made, so that it was necessary to tabulate the "work" and "study" children of all grades into one total, giving one "work" group of 134 children and one "study" group of 37 children for the three years combined. The year groups of course include the "work" and "study" children as well as the "play" children but the two former are comparatively so few in number that in any one grade they affect the results very little, so that for purposes of comparison the year group as given may be considered as a "play" group.

Results

Although we are here concerned with the influence of the vacation on the results in the various groups only, it is of interest to note the distribution in the different tests. From a study of the tables of distribution which were derived but are not given here for want of space, several things are clear. In speed there is a wide distribution in all groups, being particularly marked in the seventh year and being of wider range for the addition and subtraction tests than for the multiplication and division tests. So marked is this range of distribution that at times we find one third of the fifth year group showing a better rate than one half of the seventh year group and, conversely, about one third of the seventh year group showing a rate poorer than one half of the fifth year group. This wide range in rate exhibited by all the groups would seem to indicate the necessity for special drill for those pupils in the lower end of the scale, and for relief from drill of those children at the upper end of the scale. That this is not done more often than it is done at present is due to difficulties inherent in the crowded condition of our classes.

TABLE OF MEDIANS AND VARIATIONS FROM MEDIANS

Group	Test I			Test II			Test III			Test IV		
	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation	June Median	Sept. Median	Median Variation
RATE												
7th Year	11.1	12.7	+1.6	13.1	12.9	– .2	10.5	10.5	.0	10.1	9.9	– .2
6th Year	8.7	10.6	+1.9	11.5	11.0	– .5	9.7	8.8	– .9	8.9	8.8	– .1
5th Year	8.6	8.8	+ .2	10.4	9.6	– .8	8.2	7.9	– .3	6.5	6.4	– .1
Total "Play" Group	....	....	+3.7	....	....	–1.5	....	....	–1.2	....	...	– .4
Average "Play" Group	....	....	+1.2	....	....	– .5	....	....	– .4	....	...	– .1
"Work" Group	10 7	11.1	+ .4	12.2	11.7	– .5	10	9.7	– .3	9.6	9.0	– .6
"Study" Group	9.7	11.2	+1.5	12.0	11.3	– .7	9.4	8.8	– .6	9.0	7.9	–1.1
ACCURACY												
7th Year	81	69	–12	93	82	–11	85	73	–12	89	84	– 5
6th Year	78	71	– 7	88	87	– 1	83	74	– 9	89	83	– 6
5th Year	75	70	– 5	87	88	+ 1	81	70	–11	79	70	– 9
Total "Play" Group	..	..	–24	..	..	–11	..	..	–32	..	..	–20
Average "Play" Group	..	..	– 8	..	..	– 4	..	..	–11	..	..	– 7
"Work" Group	79	70	– 9	88	86	– 2	81	71	–10	81	81	0
"Study" Group	80	70	–10	91	86	– 5	81	72	– 9	88	84	–4

Except in the matter of distribution, the accompanying table gives a summary of all the results in terms of medians and variations of medians from June to September. A study of these enables us to summarize the results as follows:

1. In speed there is a general decrease from June to September, except in addition. This decrease, however, never amounts to more than a year's work. The "work" group as a rule presents a more favorable showing than the "study" group and as good or better than the "play" group except in multiplication.

2. In accuracy there is a decided decrease in efficiency in all groups. This decrease averages the equivalent of almost two years' work with slight exceptions in Test 2 and Test 4. The highest median in September is lower than the lowest median in June. The worst showing is made by the seventh year group which shows the highest median in all tests in June and the lowest or next to the lowest in September (except in division). The loss of efficiency in this group at times amounts to more than the equivalent of two years' work, the September median in addition, for example, being lower than the June median for the same test of the fifth year. The "work" group as a rule shows more favorable results than the "study" group.

3. When taken in conjunction with the increased rate, the decrease in accuracy in addition is significant. "Familiarity has bred contempt." Although the children have probably been drilled more often in addition than in the other fundamentals and thus have acquired a facility in handling such examples, the difficulties of the process have not been given close attention. The many processes involved, ("carrying," the "attention span," and other difficulties) have not been given the conscious attention necessary for the formation of a habit. In the fifth year, where the habit is of recent formation, the loss in accuracy is not so great and the gain in speed is negligible.

4. By comparison with the "play" group (averaged for the 5th 6th and 7th years) the "study" group shows the greatest loss in rate (except in addition, where it shows the greatest gain) while the "work" group is as good or better than the "play" group, except in division. In accuracy the "work" group shows a much smaller loss than the "play" group while the "study" group shows a greater loss in addition and subtraction and a smaller loss in multiplication and division.

Conclusions

Though the vacation brings about a general loss in efficiency both in speed and in accuracy, the loss in speed is such that it may be neglected in an attempt to make up the loss in accuracy. On return to school the emphasis, for a time at least, should be on accuracy to the exclusion of speed. In addition, particularly, accuracy should be emphasized. In all fundamentals those pupils

showing proficiency should be excused from drill whereas those showing particularly poor results should be placed in special drill classes.

The comparatively better showing of the "work" group and the poor showing of the "study" group would seem to indicate that the child who keeps his brain alert by meeting actual problems of life is in a better position to retain what he has learned than the one who plays or studies. The poor showing of the "study" group, however, may be partly explained by the fact that the studying group did not continue to the opening of school. The last four weeks of the vacation were spent in play by this group.

The great loss in efficiency of the "play" group would seem to point to the advisability of an all year school for the child. Whether this conclusion is justified, however, would depend upon the power for recovery shown by each group. This would have to be the subject of further tests.

Maurice A. Garfinkel.

Public School 9, The Bronx,
New York City.

A COMPARISON OF THE RESULTS OF GROUP TESTS AND THE POINT SCALE TEST

Every year mental measurements assume a greater importance in this country. As public school teachers come to have a scientific knowledge of applied psychology so that they can use psychological methods and understand and interpret the results of mental measurements, such measurements will come into general use. Hitherto the various adaptations of the Binet Scale have been used almost exclusively, in the practical diagnosis of subnormality. But the use of the Binet tests or any other individual tests for all the pupils of a school system is now, and perhaps always will be, an impossibility. However, careful annual testing of *all* the children in our schools is highly desirable. There is, therefore, a present demand for the extension, standardization, and perfection of group tests. If group tests give practically the same results, measure the same important functions, then there is no reason why they should not be used instead of the individual tests except in the case of the extreme variants, who should, of course, have a much more extensive examination than is possible with a few group tests.

The writers have undertaken a comparison of the results obtained by use of the Yerkes-Bridges Point Scale with the results obtained by the use of standardized group tests. The subjects were eighty public school children in grades four, five, and six. The Yerkes-Bridges tests were carefully given to the children dur-

ing the fall and winter of 1917-18. We began with the higher grade and proceeded to the lower. The group tests used were the seven mental tests standardized by one of the writers (W. H. P.).* They were administered simultaneously to all the children of each grade. The individual tests were all given, and the results of all the group tests were graded by one of the writers (G. E. B.). Absolute uniformity was therefore secured in the grading and evaluation of the results. The correlation by the Pearson formula between group test rating and Point Scale rating was .619; P. E. .046. If instead of using the absolute marks of the pupils, we use their ranks and compute the correlation by the Spearman foot rule formula, we get R = .84.

The pupils of each of the three grades were then ranked on the basis of the teachers' judgments of their ability. The correlation between group test ranking and the teachers' ranking, obtained by the use of the Spearman Foot-Rule formula, were: fourth grade, .565; fifth grade, .71; sixth grade, .80. Correlations by use of the same formula between Point Scale ranking and teachers' ranking were: fourth grade, .42; fifth grade, .95; sixth grade, .655. The average correlation, therefore, between the group test ranking and teachers' ranking was .692; and the average correlation between Point Scale ranking and teachers' ranking was .675. These correlations are practically the same, that from group tests being .017 the higher. Therefore, whatever value inheres in the Binet Scale as a means of measuring intelligence, the same value inheres in the group tests. By every criterion that it was possible to apply, group test rating was found to be as accurate as Binet test rating.

If the raw correlation by the Pearson formula is .619 and by the Spearman formula .84, the true correlation between the two sets of measurements is probably near unity. Nor should it be surprising that such is the case, for the Binet tests, as well as the group tests, are merely measures of the various aspects of learning capacity, the efficiency of the associative processes, and the effectiveness of memory, that is to say, in both cases we are probably measuring the same mental functions. The only question is as to the most accurate method of measuring these functions. Our results seem to show that the group test method is as accurate as the individual test method.

The significance of this conclusion is, we think, great. When we further perfect and standardize our group tests, the teachers of our public schools will have at their command a fairly accurate method of measuring the mental development of their pupils. By combining the results of a series of group tests—tests that can be given in about an hour—the teacher can, within a few days after

The Examination of School Children, The Macmillan Co., 1913. *Manual for the Mental and Physical Examination of School Children*, University of Missouri Extension Bulletin, No. 21, 1916. *The Science of Human Nature*, Silver, Burdett & Co., 1917.

the beginning of her term of school, have a fairly accurate measure of the native capacity of the children in her grade, and can rank them with reference to their ability as accurately as she can on the basis of a year's experience with them. Of course, we realize that a single measure of rote memory, logical memory, or learning capacity is no more—nor less—accurate than a single measure of jumping capacity, running capacity, or any other physical ability. However, if single measures of the various important mental functions are made and the results pooled, since each is to some extent a measure of general mental development or intelligence, their combined results have considerable validity. Moreover, the tests of the several mental functions can be repeated till, let us say, four of each have been given, at the expense of much less time and energy than could possibly be the case in the use of individual tests. By such repetitions we would not only have a fairly accurate measure of general intelligence by combining the results, but we would have a tolerably accurate measure of each of the several functions.

The accurate gradation of the pupils is of the highest importance not only to the school but to the tax payers who support the schools. In a recent survey of the mental development of the pupils of an entire school system (Webster Groves, Mo.) the children were found to be scattered through the grades with little reference to ability. The frequency surface of fourth grade ability was found actually to overlap to some extent the frequency surface of high school ability. From every point of view this great variation of ability in the same grade is highly undesirable. The school children should be put into small groups of fairly even ability, and each group allowed to proceed in its mastery of the work outlined in the course of study as fast as the group's ability warrants. It is not economical, even from the point of view of the tax payer, to have a pupil repeat a grade. Accurate gradation and proper methods of teaching will make such a thing as a repeater an impossibility. Pupils should not repeat grades, but pass through grades at various rates of progress depending upon ability.

Mental tests in the future will be found helpful in the work of accurate gradation on the basis of ability. This conclusion is no longer a mere theory or opinion, for numerous studies in this country and in England have shown beyond doubt the validity of mental measurements. The present study does, we hope, throw some light on the relative validity of group tests.

G. E. Breece
W. H. Pyle

University of Missouri.

ABSTRACTS AND REVIEWS

E. A. Doll. *Clinical Studies in Feeble-Mindedness.* Boston: Richard G. Badger, 1917. Pp. 232. $2.50.

For ten years Dr. H. H. Goddard with a corps of able assistants has been studying the problems of feeble-mindedness in a most careful and scientific way at the Vineland Research Laboratory. When he began his researches relatively little thought was given this subject; now, feeble-mindedness is considered a profound sociological and economical problem and, as a consequence, almost every body from women's clubs to "Uncle Sam" is talking about it. Such confidence, of late, has been placed in the students of feeble-mindedness that the educationalists are demanding of them how to classify children in school; judges to tell them what to do with offenders; eugenists what people should marry; and "Uncle Sam" has commandeered hundreds of these students to tell him whether his boys are fit to fight for freedom.

The Vineland Laboratory is responsible for initiating this great movement and it has done more real research work, collected and evaluated more data, and, in general, contributed more toward giving to students of feeble-mindedness their standing today than any other institution in the world. Therefore it is very fitting that some one representing this laboratory should tell the outside world what conclusions, in reference to feeble-mindedness, they have come to, and something of the procedure followed by them in coming to these conclusions.

This has been done, unusually well, by Mr. Edgar A. Doll in his book entitled *Clinical Studies in Feeble-Mindedness.* Mr. Doll was, for five years, assistant psychologist in the Vineland Laboratory and is now serving as a lieutenant in the department of psychology in the army.

In this book, Mr. Doll has given to the public the important findings which a trained mind can glean relative to feeble-mindedness by living with the feeble-minded and by studying them under the most favorable circumstances. The book is wonderfully condensed—intended primarily for the student of feeble-mindedness—yet it is so readable that it cannot fail to be of great interest and value to the layman. Educators, doctors, lawyers, in fact, anyone who wishes to know the present status of our knowledge relative to feeble-mindedness will find this book not only a rich store house of information, but the many explicit references given by the author make it easy for anyone to pursue the subject further if desired.

The book is divided into two parts. Part I deals with diagnostic criteria and is divided into three chapters. The first chapter sets forth in a clear cut way the problems to be considered. The scope of chapters two and three can best be judged from the following table of contents:

Chapter II, Criteria and Definitions. The Psychological Criterion. The Social Criterion. The Pedagogical Criterion. The Medical Criterion. The Somatic Criterion. The Hereditary Criterion.

Chapter III, Diagnostic Methods and Values. General Principles. The Diagnostic Syllabus; Personal History; Clinical Examination. The Family History Examination; Method; Value for Diagnosis; Value for Etiology; Types of Mating and Expected Offspring. The Social Status Examination; Method; Value. The Medical Examination; Method; Value. The Somatic Examination; Method. The Pedagogical Examination; Method; Value. The Psychological Examination; Importance; Methods; Value.

Part II is taken up with the consideration of six "Illustrative Cases."

The tests used and studies made in family history and other sources of information are recorded in full, and are accompanied by a diagnosis and a prognosis. The treatment of these cases is most illuminating. Contrary to the commonly accepted opinion that the Binet tests are about the only criterion used in passing upon a person's mentality, Mr. Doll in testing these six cases used about twenty-five other tests. He gives the reaction of each case to each test and designates its value in the final diagnosis. The following are the cases considered:

Case I. Obviously Feeble-Minded by Binet-Simon Tests.

Case II. Potentially Feeble-Minded.

Case III. Potentially Normal.

Case IV. Deaf-mute, Neurotic or Psychopathic but not Feeble-minded.

Case V. Psychopathic with slight Intellectual Retardation, not amounting to Feeble-Mindedness.

Case VI. High Grade Borderline Immigrant.

These cases are not intended to be typical but the method of arriving at a diagnosis in each case is typical of the method of procedure that should be followed before making a final diagnosis. The book closes with a carefully selected bibliography of seventeen pages followed by a glossary of the more common technical terms used in the text, and a brief description of thirty tests often used in making mental examinations.

J. M. McCallie.

Supervisor Special Classes,
Trenton, N. J.

A. J. ROSANOFF, HELEN MARTIN, AND ISABEL R. ROSANOFF. *A Higher Scale of Mental Measurement and Its Application to Cases of Insanity.* Psychological Monographs, Vol. 25, No. 3, Whole No. 109. Princeton, N. J.: Psychological Review Company, 1918. Pp. 113. $1.50.

The authors have devised a test on the order of the well-known Kent-Rosanoff association test, but in which the list of 100 stimulus words is drawn from technical terms such as would be encountered in high school and college studies. They have then obtained the simple verbal reactions to these words of 1000 persons of at least a full collegiate education and of 100 noted men of science (many of them in the starred list in American Men of Science). In use the value of each of the 100 responses of a given subject is obtained by reference to frequency tables based on these 1000 or 100 persons and the sum of these values is compared with the average sum for the persons of collegiate training or preferably with the average sum for the noted men (which is referred to as the "high standard" value.) Norms are given for persons of elementary school or high school education and a report is made of the responses of 205 neuropathic individuals. The outcome of the test is influenced not only by education, but also by physiological maturation, sex and temporary condition of the examinee. Without seeking to enter upon a critique of this investigation, the question may be raised whether the method here proposed is not a clumsy, lengthy and confessedly unreliable method of arriving at information that can be gotten by various tests already in the market much more precisely and in a fraction of the time.

G. M. WHIPPLE.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except July and August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated.

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of ten monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X JANUARY, 1919 NUMBER 1

EDITORIAL

MENTAL MEASUREMENTS IN EDUCATION

President Wilson has said that the war has brought forth some very terrible things, but it has also given rise to some very beneficent and valuable things. He was speaking particularly of the conditions in Europe, but the statement applies to this country as well. One of the incidental by-products of our military activity has been the stimulation of popular interest in mental measurements. When the war broke out the psychologists immediately mobilized and offered their services to the government. Not only that but they devised and induced the government to adopt a plan for the testing of recruits that was unlike anything that the world had ever seen before. It is too soon to say just how much this psychological testing and the utilization of the results contributed to increase our military efficiency. Perhaps we shall never be able to tell. But the mere fact that recruits were given mental tests, that in such a life and death matter as waging war the laboratory psychologist had something to offer that was of practical significance, caught the popular fancy and stirred the im-

agination of the man in the street. Never before has the recondite subject of psychology been held in such high esteem. Never before has the daily press devoted so much space to the possibilities of mental measurements. Never before has there been such a favorable opportunity to extend the use of mental measurements in education.

In a circular letter to city superintendents Commissioner Claxton adverts to the fact that many cities have already established bureaus of educational research, and asserts that every city of more than 30,000 inhabitants should have such a bureau. He calls the attention of school authorities to the two or three hundred young men who have been working in the psychological division of the army, and are now about to be discharged, and stresses the "unusual opportunity for city schools to obtain the services of competent men as directors of departments of psychology and efficiency, for such purposes as measuring the results of teaching and establishing standards to be attained in the several school studies, applying mental tests and discovering mental aptitudes of pupils, discovering defective children and children of superior intelligence, and investigating various other vital questions necessary to establish an intelligent basis for promotions, class organization, and special schools."

The suggestion is an excellent one, and it is to be hoped that many school systems will heed the call, and will establish bureaus or enlarge those already established. The time is indeed opportune to press the matter of educational research in each community, and the popularization of the role of psychology in the war tends to create a favorable attitude to such work on the part of the public. But even before the war there was no dearth of trained men and women in our schools capable of directing such bureaus and carrying on such research. The trouble has been that superintendents, boards of education, and the public at large have not been sufficiently convinced of the practical value of mental measurements in education to support them financially and to assign competent teachers to conduct them. It is now incumbent upon all those who believe in the value of scientific studies in education to engage in a vigorous and aggressive campaign to arouse public opinion to demand such studies in the schools as a matter of enlightened and progressive educational policy.

J. C. B.

NOTES AND NEWS

At the December meeting of the New York Society for the Experimental Study of Education the general topic for consideration was "Problems in the Teaching of Civics." Associate Superintendent John L. Tildsley spoke on "Some Needs in Civics Teaching," stressing the need for better teachers, for more time devoted to the subject, and for a more vital connection with the life activities of young people. Professor David Snedden, of Teachers College, in "Practical Problems in Civics" presented two problems for experimental investigation: First, in what respects are selected groups of citizens now failing in citizenship, and what means of education can be devised to supply these defects; second, to what extent can and should the civic virtues of the school community be carried over to the civic community? Miss Mabel Skinner, of the Washington Irving High School, taking as her topic "Community Civics: What Next?" contended that civics should be treated as a laboratory subject, and described the experiments she had been conducting. Mr. Eugene B. Gartlan, principal of Public School 42, The Bronx, discussed "Civics Teaching in Evening Schools," and narrated the difficulties of teaching civics to Russian Bolsheviks in New York City.

At the University of Cincinnati Professor Ernest L. Talbert is giving a course of lectures on *Race Psychology*. Professor Guy A. Tawney, one on *The Nations at War* and Professor Harris M. Benedict, one on *Evolution—Its Nature and Effect upon Human Ideals.*

At the Baltimore meeting of the American Psychological Association Major Bird T. Baldwin presented a paper on *The Function of Psychology in the Rehabilitation of the Disabled Soldier* in which he discussed the aim and scope of psychological service and reported detailed studies in the psychology of voluntary movement. The paper was illustrated with lantern slides.

Professor Samuel C. Kohs, of Reed College, Portland, Oregon, has devised an ingenious instrument for determining the Intelligence Quotient in cases of mental measurement. He calls it the Reed Intelligence Quotient Slide Rule. It consists of a movable cardboard disc with a chronological age scale on its edge centered over a circle with a mental age on its inside and an intelligence quotient scale on its outside. A celluloid indicator is set at the appropriate mental age, the central disc is rotated until the chronological age falls under the hair line, the indicator is then moved to the "I. Q. Line" and the Intelligence Quotient is read directly from the outside scale. With practice any I. Q. can be determined within 15 seconds, and 100 of them read in 25 minutes. The slide rule may be procured from the Reed College Cooperative Store, postage paid, for thirty-five cents.

District Superintendent Arthur C. Perry, Jr., of New York City, recently gave twenty questions on the War to one thousand pupils of the eighth-grade graduating classes. The results were as follows:

1. What do the letters "W. S. S." stand for? 99.5%
2. Who is the commander of the American forces in France? 94.2%
3. Name an illegal use Germany made of her submarines 89.9%

4. In what month and year did the U. S. enter the war? 89.2%
5. Name a battle in which U. S. troops took part 84.4%
6. Upon what nation did Austria first declare war? 80.6%
7. What nation was the first to declare war? 78.3%
8. What do the letters "A. E. F." stand for? 73.5%
9. Name three European nations which were not in the war 70.4%
10. With what nation did Germany make peace in February, 1918? 70.1%
11. In what battle in the first year of the war were the Germans stopped in their advance on Paris? 66.4%
12. With what nation did Germany intrigue, promising to help it invade the United States? 63.4%
13. By what nation were the German colonies conquered? 61.7%
14. In what continent were the chief German colonies? 59.4%
15. About how many soldiers did the United States send to Europe? 58.5%
16. Name a legal use a nation may make of a submarine in war 58.2%
17. In what month and in what year did Austria declare war? 40.9%
18. In what great naval battle were the Germans repulsed? 38.6%
19. Italy had an agreement to fight with Germany and Austria. Why did she fight against them? 36.7%
20. With what nations is the United States at War? 31.2%

Among the sub-committees appointed by the Commission on Secondary Education of the North Central Association is one on Educational and Vocational Guidance, the chairman of which is Professor Guy M. Whipple, of the Carnegie Institute of Technology.

During the illness of Professor Herman Harrell Horne, of New York University, three of his classes are being conducted by Dr. J. Carleton Bell, of the Brooklyn Training School for Teachers.

At the recent scientific meetings in Baltimore Colonel Walter D. Scott was elected president of the American Psychological Association, and Major Robert M. Yerkes vice-president and chairman of Section H (Anthropology and Psychology), of the American Association for the Advancement of Science.

Professor William Harry Heck, of the University of Virginia, died of pneumonia following influenza, January 4, 1919, at the age of 39 years. Professor Heck contributed extensively to the experimental study of education, and his book on *Mental Discipline and Educational Values* is an excellent summary of the arguments on this question. He published many papers on *Mental Fatigue in Relation to School Work*, and at the time of his death was preparing an *International Source-Book on Home Study*.

Mrs. Maria Kraus-Boelte, well known for her manifold activities in the spread of the kindergarten movement, has died in New York City at the age of eighty-two years.

Dr. C. Ward Crampton, director of physical training in the schools of New York City, has been granted a year's leave of absence with half salary by the Board of Education for the restoration of his health.

Dr. Raymond Dodge, professor of psychology in Wesleyan University, has been made lieutenant commander in the United States Navy, in recognition of the special psychological tests devised by him for the use of the Navy during the war.

PUBLICATIONS RECEIVED

ROSE LUCIA. *Peter and Polly in Autumn.* Cincinnati: The American Book Company, 1918. Pp. 176. Forty-eight cents.

A delightful story that makes a strong appeal to the little folks. Excellent supplementary reading for the second and third grades.

HELEN MAROT. *Creative Impulse in Industry.* New York: E. P. Dutton and Company, 1918. Pp. xxii, 146. $1.50.

This is a book which challenges the thoughtful attention of every one interested in industrial education. It is characterized by ideas, by grasp of fundamental principles, by earnestness and enthusiasm, by a vision of revolutionary possibilities in the relation of education to industry. In the period of history upon which we are now entering industrial questions are certain to be of the profoundest importance for social welfare. Is the statement made by employers true that from 95 to 99 per cent. of the working force is without productive impulse? If it is now true, is it due to the limitations of human nature and the unalterable demands of modern industry, or is it due to faulty organizations of industrial enterprises? If the former is true, then there is little hope for an industrial democracy. But the author holds that the latter is the real explanation, and looks to education to provide the stimulus for the creative impulse in industry. In the final chapter an extremely interesting experiment in industrial education is outlined, and we are given to understand that there is a probability that the project will be given a fair trial. If so, it will be one of the most valuable and significant contributions that has yet been made to the subject of industrial education.

EUGENE A. NIFENECKER. *Measurements in Spelling* New York: Department of Education, Division of Reference and Research, Publication No. 16, 1918. Pp. 88.

This valuable study of spelling was undertaken at the request of District Superintendent James J Reynolds, in charge of the schools on Staten Island. Twenty-five words from Group T of the Ayres Scale were spelled by each of 5260 pupils in Grades 5A to 8B inclusive. Each word thus received almost four times as many spellings as in Ayres' original study. All grades except the fifth attained higher percentages than indicated on the Scale. The results are tabulated by schools and grades, the age-grade relations are plotted, sex differences are brought out, the home language factor considered, and the factors of time allotment, size of class, and departmental teaching are carefully weighed. A valuable phase of the study is the appendix giving the proportion of correct spellings on each word for each sex of each half grade separately.

NAOMI NORSWORTHY AND MARY THEODORA WHITLEY. *The Psychology of Childhood.* New York: The Macmillan Company, 1918. Pp. xix, 375. $1 60.

This book is intended as a text-book in normal schools. We are told that a course in general psychology is presupposed, though it is difficult to see why any intelligent student might not cope successfully with the text without such preliminary study. Each chapter is followed by a list of exercises, questions for discussion, and references for futher reading. "Constant emphasis has been thrown on the

physiological basis of the tendencies discussed, and Thorndike's classification of instincts, on the basis of the responses made, is adhered to throughout. The greatest space is devoted to a descriptive study of children as differentiated from adults." The chapter headings reveal something of the scope of the work. They are: The source of original nature; characteristics of original nature; tendencies resulting in action, non-social instincts; social instincts; tendencies accompanied by affective states; attention; sense perception; memory; imagination; thinking; habit and learning; play; moral and religious development; physical development of the child; a cross-section of child life at five and at eleven; exceptional children; and methods used in psychology. It will be seen that the authors follow fairly closely the rubrics of traditional psychology, yet examination of the text shows a wide divergence from traditional treatment. Throughout pains are taken to select from the rich literature of the topics treated those data that pertain specifically to the development of childhood. It is a distinct advance over any child psychology that we have seen.

Paul Popenoe and Roswell Hill Johnson. *Applied Eugenics.* New York: The Macmillan Company, 1918. Pp. xii, 459. $2.10.

This is a fascinating book on a subject of increasing social importance. The authors, one the editor of the *Journal of Heredity,* the other professor of sociology in the University of Pittsburgh, are admirably equipped to give a presentation of the subject both scientific and interesting. Their success has been marked. It is safe to predict that the book will speedily be recognized as the chief work on applied eugenics in any language. May it be widely read in order to counteract the many absurd beliefs that are held by otherwise intelligent people! One of the most interesting portions of the book is the chapter on The Eugenic Aspect of some Specific Reforms, including taxation, back to the farm, democracy, socialism, child labor, compulsory education, vocational guidance, minimum wage, mothers' pensions, housing, feminism, old age pensions, sex hygiene movement, trades unionism, prohibition, and pedagogical celibacy. Other important chapters are those on religion and eugenics, and eugenics and euthenics. Almost every chapter has something of significance for the student of education.

Mabel Powers. *Stories the Iroquois Tell Their Children.* Cincinnati: The American Book Company, 1918. Pp. 216.

Nothing is more indicative of the temperament of a people than the stories which it tells to its children. The stories in this volume have been gathered at first hand from the surviving Indians in New York State, and their fidelity is attested by the chiefs of the tribes. They are interesting folk tales which gain an added attractiveness from the numerous illustrations, many of which are in color. The book is well adapted for supplementary reading in the intermediate grades.

Kenneth Richmond. *The Permanent Values in Education.* New York: E. P. Dutton and Company, 1917. Pp. xxiii, 136. $1.25.

This little book consists of a series of excursions into various fields of the history of education to discover and bring forth the lessons which educational movements of other days have for our own time Examination of the educational systems of the Jews, the Greeks, the Romans, the Renaissance, and the theories of Milton, Comenius, Locke, Rousseau, Pestalozzi, Herbart and Froebel, reveals certain funda-

mental principles which are valid for all time. The chief of these is that education must aim to make every man and every woman of a nation able to enjoy the entire fulness of life which the nation offers.

Schoolmen's Week Proceedings at the University of Pennsylvania. University of Pennsylvania Bulletin, Series 18, No. 5, Part 3, 1918. Pp. 362.

This volume contains many papers that are of interest to the scientific student of education. Among these are three addresses by Professor Charles H. Judd, of the University of Chicago, on "Educational Measurements—Their Benefits to Teachers and to Laymen," "Measurement of Reading," and "Measurement and the Promotion of Pupils;" "Report on Co-operative Work in Educational Measurements," by Harlan Updegraph and LeRoy A. King; "The Use of Educational Measurements in Formulating Changes in the Course of Study," by J. N. Adee; "Measuring the Fundamental Operations of Arithmetic in the Scranton Schools", by Thomas Francis; "The Courtis Tests in Arithmetic in a Group of Philadelphia Schools," by Philip A. Boyer; "A Study of Individual Pupils with the Aid of Courtis Tests in Arithmetic," by J. M. Fisher; "The Use of the Courtis Test in Silent Reading," by Katherine E. Moran; and "A Study in the Use of the Courtis Arithmetic Tests in Lock Haven," by C. W. Hunt.

F. L. and A. M. Skinner. *Happy Tales for Story Time.* New York: The American Book Company, 1918. Pp. 180. $0.64.

These delightful stories will provide many happy hours for tired and restless children. There are groups of animal tales, grandmother's tales, tales of Christmas, folk tales and fables, and wonder tales. The last story in the book is the Blue Bird, adapted from Maeterlinck.

George D. Strayer and Frank P. Bachman. *The Gary Public Schools, Organization and Administration.* New York: General Education Board, 1918. Pp. xix, 126. $0.15.

This is the second volume of the Gary Survey. The introduction gives an outline of the Gary Plan and this is followed by a chapter setting forth present day problems, such as modernizing the curriculum, increasing the efficiency of instruction, and financing the educational enterprise. There follow chapters on program plant, organization, use of plant, supervision and administration, and comparative cost. The general supervision is characterized as inadequate in amount and ineffective in its results. The principals are administrative officers who have little to do with the educational aspects of the schools. The administration has new and difficult problems to meet, and has not shown great efficiency in their solution. This seems to be one of the weakest points in the Gary regime.

C. W. Sutton. *The Results of a Spelling Test.* Cleveland: Board of Education, Division of Reference and Research, 1918. Pp. 19.

Lists of words were selected from the Ayres Spelling Scale and given to each grade from III to VIII. Each list contained from 30 to 50 words. and was composed of different words for each grade. This makes comparison of grades impossible. There were 1,941,185 spellings obtained from 50,687 children. A valuable feature of the record is the percentage of correct spellings on each word for each grade. The results of two years' drill upon the Ayres list of 1000 words are to make the records of the different schools much more uniform, and to change somewhat the relative position of the words in the scale of correct spellings.

VOLUME X FEBRUARY, 1919 NUMBER 2

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE EFFECT OF PRACTICE UPON VISUAL APPREHENSION IN THE FEEBLE-MINDED[1]

KARL M. DALLENBACH, PH. D.

Cornell University

I. INTRODUCTION

In the experiment on the effect of practice in visual apprehension which we performed upon second grade school children,[2] we learned that the group classified as "poor," and composed, upon the basis of the performance in the first week's work, of the lower third of the class, made a slow and continued improvement; whereas, the groups classified as "medium" and "superior" made, what we have learned to expect in practice experiments, a large improvement in the earlier and a small improvement in the later stages. The "poor" group ultimately came to surpass the record of the "medium" group, and to approach that of the "superior."

Problem.—Since the "poor" group was composed of the most backward children in the class, most backward not only in respect to their performance in the tests of visual apprehension, but also in respect to class standing and teacher's estimate,[3] it was deemed advisable to extend the work to a class of feeble-minded children in order to discover whether a continued improvement in visual apprehension is characteristic of backwardness, and whether the effects of practice are as great as in the case of the lowest third of a normal group.

As in the earlier experiment, our results are purely quantitative, and we refer for their interpretation to the introspective studies that have been made upon this subject.[5]

[1]The experimental work was performed at the Ohio State University.

[2]DALLENBACH, K. M., *The Effect of Practice upon Visual Apprehension in School Children.* THIS JOURNAL, V, 1914, pp. 321–334; 387–404.

[3]*Op. cit.*, pp. 388, 400, 401.

[5]WHIPPLE, G M., *The Effect of Practice upon the Range of Visual Attention and Visual Apprehension.* THIS JOURNAL, I, 1910, pp. 249–262.

FOSTER, W. S., *The Effect of Practice upon Visualizing and upon the Reproduction of Visual Impressions.* THIS JOURNAL, II, 1911, pp. 11–22.

Subjects.—The subjects were inmates of the State Institution for the Feebleminded at Columbus, Ohio.[6] Forty-two children, nineteen girls and twenty-three boys, took the tests. We found, however, when it came to the grading of the results that the records of one girl, who was partially paralyzed, had to be discarded as they were illegible.

An effort was made when we were selecting our subjects to choose children who would compare mentally with those who served in the previous experiment. Unfortunately no mental tests had been given in the earlier work, and we had no definite knowledge of the subjects' mental ages; but as the drill was given in the second grade during the second semester we felt safe in assuming that the average mental age was between 7 and 8 years. Accordingly, subjects were selected from the school for feebleminded whose mental ages at the time of their admission to the institution, as determined by the Binet-Simon tests, averaged 7.5 years, with a mean variation of 0.58 years. The chronological age ranged from 10 to 18.

II. METHOD AND RESULTS

A. Method of Procedure

1. Material.—The same kind of material was used as in the earlier work,[7] namely: I. numerals; II. letters; III. combined numerals and letters; IV words; V. combined numerals, letters, and words; VI. geometrical figures; and VII various materials, numbers, letters, figures, signs, etc., arranged in a "tit-tat-to" frame. These seven types were each divided into a number of sub-types of greater and lesser difficulty.

In the first five types, the material on the one half of the cards appeared in a vertical column, and on the other half in a horizontal row. All the material was printed by hand in India ink on white card boards 11x14 inches in size. All letters, numerals, and other characters upon the cardboard were 2 inches in height. The letters and numerals were printed in script and block; the letters in both upper and lower case.

2. Plan of Experiment.—Exactly the same procedure was followed as in the earlier experiment. The work was divided into nine periods—a preliminary practice, five test, and three drill periods. The experiment began on the 21st of February, 1916, and ran to the 2nd of March, 1917.

---

[6]The writer takes this opportunity to thank Dr. E. J. Emerick, superintendent of the institution, for the privilege of conducting the experiment and also for permission to use the institutional records.

[7]Pp., 326–330.

The plan of the experiment in outline, is as follows:

WEEK	WORK	DATE
1	Preliminary Practice Series	February 21–February 25
2	Test Series A	February 28–March 3
3–6	Drill Series 1	March 6–March 31
7	Test Series B	April 3–April 7
8–11	Drill Series 2	April 10–May 5
12	Test Series C	May 8–May 12
13–16	Drill Series 3	May 15–June 9
17	Test Series D	June 12–June 16
53	Test Series E	February 26–March 2

The material presented in the *Preliminary Practice Series* and the five *Test Series* was an exact duplicate of that given to the normal children, and was printed in block and in small letters. The *Preliminary Practice Series* was composed of the easiest types of material. We hoped by means of it, not only to familiarize the subjects with the experimental method, but also to obtain a basis for classification into poor, medium, and superior groups. The *Test Series* was composed of difficult types of material, and consisted of two sets of test cards. One set was used for Test Series A, C, and E, given 9 and 41 weeks apart respectively; and the other set for *Test Series B* and *D*, given 9 weeks apart. The geometrical figures which appeared in the first set of test cards were revolved through an angle of 45° for *C*, and 90° for *E*; and those which appeared in the second set were revolved through an angle of 45° for *D*. The second set of test cards had exactly the same number of elements and the same materials as the first. It was formed by the inversion of the material in the first set. With the materials of the first five types, the order was merely reversed: what was at the head of the line was now placed at the end; what was at the top of the column was now placed at the bottom. The geometrical figures, Type VI, were revolved through an angle of 180.° The same characters were used in Type VII, their relative positions in the "tit-tat-to" frame were, however, changed.

The *Drill Series* differed radically from those of the earlier experiment. They were more systematically planned and presented, and they were printed, as was not true in the other experiment, in both block and script style, and in both capital and small letters.

Drill Series 1 was composed of comparatively simple materials. During the first week, cards were used, the elements of which—counting every numeral, letter, word, line, and character an element—totalled 55 daily; during the second week 57; the third week 59; and the fourth week 63 daily. In *Drill Series 2* the material was gradually made more complex, totalling 66 elements per day for the first week; 68 for the second; 71 for the third; and 75 for the fourth week. This progression toward complexity was continued in *Drill Series 3* in which the first week's work totalled daily 80; the second, 85; the third, 90; and the last, 100. During the last week the most difficult cards from each type of material were selected. The plan of the series may be summarized as follows:

Drill Series	Week			
	1st	2nd	3rd	4th
1	55	57	59	63
2	66	68	71	75
3	80	85	90	100

The girls and boys, as required by institutional discipline, were given the experiments separately; the boys, the first 15 minutes in the morning; and the girls, the first 15 minutes in the afternoon. In both cases the experiments were conducted by the principal of the school, Mrs. Leal, with the assistance of the teachers.[8] The principal stood directly in front of the class and displayed the cards by hand. After a preparatory signal, "Ready, now," the card was exposed by simply turning its face to the class. The teacher stood at the rear of the room, and, with the aid of a stop-watch, designated by a silent signal when the exposure was to be made and when ended. The exposure lasted for five seconds, and immediately after, the subjects individually reproduced upon prepared paper the card that had just been displayed.

A series consisting of 10 cards was shown every day; five of these series were given every week for 18 weeks: no interruptions occurred. There were then, 90 series in all.

As in the previous experiment, the children were not permitted during the exposure to pronounce the materials aloud.[9] There

[8]It is a pleasant duty to acknowledge the debt that I owe Mrs. Kezia A. Leal and the teachers, Miss Josephine Weldon and Miss Harriette Zurmehley, and to express my thanks and appreciation for their interest in the problem. I also wish to thank Dr. J. E. Evans of the department of psychology of the Ohio State University for conducting the experiment for me during *Test Series E*.

[9]Pp., 331.

was a strong incentive at first to do this, but by strict supervision it was soon overcome. Before exposing materials of type VII, the children were instructed to draw the "tit-tat-to" frame in readiness for report.

3. *Method of Grading Data.*—The method of scoring was exactly the same as that employed in the earlier work.[10] For every element,—figure, letter, word, line or character,—correctly reproduced one unit was credited; for every error of transposition or insertion 0.5 was deducted except in case of words of 3 and 4 letters, when 0.3 and 0.2, respectively, were subtracted.

[10]Pp., 333–334.

TABLE I

Classification, Chronological Age, Mental Age, Coefficient of Intellectual Ability, and Net Scores by Series

Subject	Chronological Age	Mental Age B-S	Mental Age 1st Y-B	Mental Age 2nd Y-B	Coefficient of intellectual ability 1st	Coefficient of intellectual ability 2nd	Average of Preliminary Practice 1st week
Mary K.	15.9	8.2	10.40	11.40	.76	.85	51.54
Beatrice	13.10	8.4	10.80	11.20	.80	.84	51.00
Don	11.2	6.6	9.42	9.00	.83	.76	50.08
Ruth	15.2	7.4	9.42	9.71	.68	.71	49.86
Violet	13.6	7.6	9.85	11.10	.75	.84	49.86
William B.	14.0	8.2	10.60	10.80	.79	.80	49.40
Paul B.	11.0	7.8	8.68	9.14	.75	.80	48.86
John G.	12.10	7.6	9.57	11.00	.74	.84	48.34
Oliver	14.7	8.4	8.39	8.76	.55	.61	47.52
Henrietta	14.0	7.4	11.60	12.00	.89	.94	47.36
William W.	11.11	6.8	8.92	11.30	.69	.91	47.32
Pearl S.	16.7	8.2	10.20	10.80	.75	.79	46.36
Willie K.	13.0	7.2	8.39	9.00	.58	.68	46.78
Charles S.	15.1	8.4	9.28	10.20	67	.75	45.36
Frances	15.9	8.8	9.28	11.00	.67	.80	45.08
Mary M.	15.11	8.0	7.80	9.57	.47	.70	44.92
DeWitt	13.2	7.0	8.00	8.07	.51	.73	44.80
Audrey A.	15.10	8.2	8.61	11.10	.59	.81	44.02
Daniel	12.3	7.4	8.61	9.14	.63	.70	43.48
Clara	18.8	7.8	8.68	8.76	.56	.57	42.56
Edna	14.3	7.4	8.61	10.40	.60	.77	40.44
Andrew	11.10	7.8	8.84	9.00	.68	.71	40.40
Hazel	16.10	6.8	8.07	8.31	.50	52	39.96
Gertrude	15.10	9.0	9.42	10.60	.68	.77	38.96
Walter	15.11	8.4	9.14	8.92	.66	.64	38.68
Kasmer	11.0	7.0	8.31	8.92	.67	.77	38.52
Louis	13.0	6.6	9.57	9.85	.73	.76	38.00
Julia	16.6	8.8	8.68	11.20	.60	.82	37.94
Edith	12.1	7.4	8.23	8.31	.56	.58	37.86
Sadie	15.1	7.6	9.42	10.00	.69	.74	37.44
Ralph	13.4	6.8	8.07	8.23	.51	.53	37.30
Leo	14.10	7.4	8.53	8.68	.58	.60	36.84
Freddie	15.0	6.4	8.00	8.00	.49	.49	35.84
Paul L.	10.0	7.0	7.40	8.00	.61	.65	35.34
Pearl J.	17.0	6.8	8.39	9.42	.52	.65	35.20
George H.	14.5	6.8	8.39	8.46	.56	.57	33.10
George E.	15.9	7.2	7.60	8.23	.46	.52	32.84
Jennie	18.2	7.2	8.39	9.42	.52	.65	32.70
John H.	13.0	8.2	8.15	8.23	.54	.55	31.98
William K.	12.1	6.4	7.80	8.46	.51	.60	31.42
Samuel	10.8	6.0	8.07	8.84	.63	.78	28.28

Av. Test Series A 2nd wk.	Av. Drill Series 1 3rd–6th wk.	Av. Test Series B 7th wk.	Av. Drill Series 2 8th–11th wk.	Av. Test Series C 12th wk.	Av. Drill Series C 13th–16th wk.	Av. Test Series D 17th wk.	Av. Drill Series E 53rd wk.
41.90	49.10	50.08	56.25	49.82	57.50	53.10	50.16
40.68	47.58	44.82	53.55	50.88	57.23	50.72	50.86
41.46	47.20	45.24	51.30	50.24	53.01	50.82	
41.24	46.48	45.12	51.53	44.08	46.38	45.30	43.94
46.40	49.50	49.76	52.03	51.32	58.61	54.60	53.12
52.60	53.48	52.16	56.10	51.62	53.18	53.66	51.36
41.72	48.89	45.76	51.89	45.80	46.27	48.70	50.50
44.00	45.37	39.33	49.45	43.83	45.34	45.03	47.20
42.47	45.90	47.08	52.75	47.17	52.66	47.25	
40.10	46.10	45.02	51.02	49.26	53.26	50.42	52.46
44.86	48.60	45.48	50.73	47.10	52.01	50.00	47.10
40.16	45.55	45.24	48.52	49.38	48.12	49.70	48.10
43.46	47.10	46.30	52.23	48.62	49.53	49.00	49.06
43.82	48.34	46.90	51.27	48.94	52.66	46.90	48.04
40.29	45.43	42.86	47.66	45.66	53.47	47.60	47.64
34.74	42.94	44.32	46.93	46.66	40.87	44.06	43.30
40.36	46.43	43.20	50.80	45.26	46.32	43.96	39.70
39.00	42.00	40.78	41.16	40.88	41.15	39.50	42.68
43.92	45.02	41.40	43.61	41.60	41.68	44.24	43.66
36.12	41.58	37.78	45.76	41.46	49.09	40.66	38.82
38.72	42.60	40.15	44.97	42.60	42.16	44.58	43.52
37.22	40.27	41.40	49.65	36.75	40.42	37.50	37.32
35.62	38.75	37.12	43.00	38.12	42.77	39.36	33.44
41.50	46.30	43.64	47.17	42.90	48.74	40.68	43.40
39.30	40.68	37.78	43.71	43.20	46.69	43.68	43.48
37.70	42.30	40.46	45.07	45.38	48.26	45.62	42.52
36.92	41.47	39.60	38.63	41.34	39.36	35.08	38.80
40.20	45.93	47.18	49.20	49.76	52.39	51.18	47.40
33.96	42.60	37.00	47.49	49.92	45.29	43.96	41.50
35.06	39.28	41.78	45.18	42.82	44.49	45.62	44.60
32.80	31.97	31.08	32.47	34.38	39.47	35.18	36.42
36.70	35.34	34.80	38.30	34.20	36.22	37.32	
38.96	43.33	39.75	46.21	42.78	46.13	43.68	37.88
30.00	32.74	26.64	32.48	32.88	34.77	28.00	30.18
28.62	36.54	26.26	35.18	30.30	34.45	28.50	30.80
31.72	35.13	32.57	38.24	39.25	37.86	40.65	29.32
28.68	32.33	33.52	32.80	28.78	31.89	34.12	30.76
32.62	37.42	33.36	39.36	38.68	36.59	39.60	36.02
32.92	38.05	29.74	37.11	33.26	37.46	35.10	31.40
28.12	40.13	35.70	39.30	35.57	36.00	36.15	36.06
30.60	31.50	31.72	30.25	30.40	27.96	32.96	31.16

II. RESULTS

1. Principal Results:

a. Classification of Subjects: As in the earlier work,[11] the subjects were classified into three groups, Classes I, II, and III, accordingly as their records during the *Preliminary Practice Series* were good, medium, or poor. The subjects, their classification, chronological ages, mental ages, mental coefficients, and records in the different series are shown in Table I above. This Table corresponds to Table I of the earlier work.[12] A comparison of these two tables reveals the fact that the abnormal children are superior in the *Preliminary Practice Series* and in *Test Series A* to the normal children. This is also very clearly shown in Tables II and III where the class averages for the normal and abnormal appear. Here it will be seen that the abnormal are not only superior as a group, but also class for class.

TABLE II

Average and Mean Variation for Each Class of the Normal and Abnormal in the Preliminary Practice Series

Class	Normal		Abnormal	
	Average	m. v.	Average	m. v.
I	42.72	2.32	48.54	1.54
II	35.32	1.36	41.26	2.46
III	28.56	1.76	34.31	2.40
Total	35.77	5.02	41.54	5.52
Girls	34.26	3.86	42.94	4.82
Boys	36.70	5.00	40 45	5.86

(Note: The totals in this and the following tables are not the averages of averages, but final averages from the original data.)

This is extremely surprising for, as we have noted above, care was taken to choose subjects whose mental ages were presumably comparable to those of the earlier work. Since the tests upon which the mentality of the abnormal children was based were made some three or four years previously, we re-tested[13] our group with

[11]Pp. 334, 388.

[12]Pp. 388.

[13]We are indebted to Dr. Rudolf Pintner of Ohio State University and his class in mental tests for their assistance in this work.

the Yerkes-Bridges point scale in order to discover if they were now of the mental ages that the institutional records indicated.

We found that the abnormal children had advanced in their mental status during the interim since their admission into the institution, and that they were now mentally superior to the group of normal children. This increase and superiority was, as is shown in Table III, a little over a year.

TABLE III

The Average Chronological Age and Mean Variation for Each Class of Subjects for Both the Normal and Abnormal; and the Mental Ages for Each Class of Abnormal Subjects together with their Coefficients of Intellectual Ability

	Normal			Abnormal							
	Chronological Age		Mental Age	Chronological Age		Binet-Simon Test		Yerkes-Bridges Point Scale		Coefficient Intellectual Ability	
Class	Av.	m. v.		Av.	m. v.	Av.	m. v.	Av.	m. v.	Av.	m. v.
I.......	8.65	0.58	No test made, but assumed, since children were normal and in second semester of second grade, that ages between 7-8	13.74	1.43	7.73	0.47	9.68	0.76	0.730	0.069
II......	8.54	0.92		14.76	1.72	7.78	0.65	8.69	0.40	0.610	0.065
III.....	7.83	0.81		13.95	1.93	7.01	0.43	8.17	0.33	0.552	0.045
Total...	8.36	0.81		14.16	1.72	7.51	0.58	8.86	0.67	0.637	0.085
Girls....	8.07	0.83		15.59	1.34	7.83	0.51	9.21	0.82	0.652	0.098
Boys....	8.54	0.80		13.03	1.36	7.27	0.61	8.59	0.58	0.626	0.087

The average mental age of the abnormal children was, as shown by the Binet-Simon tests, 7.5,—and this was the age that we desired; but as obtained by the Yerkes-Bridges tests it was 8.8. This discrepancy explains in part the initial superiority of the abnormal subjects.

b. The Effect of Practice in the Drill Series. Although the feebleminded children showed superiority in the *Preliminary Practice Series* and *Test Series A*, they nevertheless revealed their inferiority by their reports in the *Drill Series*, and in the other *Test Series*, as well as by the character of their practice curve. These results appear in Tables IV and VI.

TABLE IV

Average Scores of the Abnormals in the Drill Series

Class	Drill Series 1		Drill Series 2		Drill Series 3	
	Av.	m. v.	Av.	m. v.	Av.	m. v.
I..........	47.799	1.587	52.044	1.529	51.840	3.280
II.........	42.978	2 031	45.522	2.644	45.240	4.032
II	36.608	3.247	38.628	4.353	37.583	3.896
Total......	42.604	4.382	45 373	5.702	45.063	5.661
Girls.......	43.648	3.236	46.997	3.826	47.364	5.791
Boys.......	41.787	5 260	44.152	6.749	43.233	6.297

In the Drill Series, the results of which appear in Table IV, there is, after an initial improvement, an actual decrease in the average number of elements correctly reported. This loss, though slight, is uniform for the group, for the different classes, and for the boys. It can readily be explained by the fact that more difficult material was presented during the last series than in either of the preceding ones. As we have shown above there were from 14 to 25 more elements presented each day in *Series 3* than in *Series 2*, and from 25 to 37 more than in *Series 1*. This, to be sure, offered greater opportunities for larger scores, but these opportunities were for the most part outweighed by the disadvantage of greater complexity.

When the results were computed separately according to sex, it was discovered that the boys alone were responsible for the decrease in *Series 3*. The girls had made an improvement in this series; not as great, however, as the loss of the boys, so that their gain was masked in the general results. Upon the basis of the explanation that we have offered above for the loss in this series, the girls apparently were not confused by the greater complexity of the material, or, if they were, not sufficiently to overcome the advantage of greater opportunities for apprehension.

We should be tempted to explain this difference upon the ground that the girls were, as is shown in Table III, mentally superior to the boys, were it not for the fact, as is shown in Table V, which is a reproduction of Table III of the earlier work,[14] that

[14]Pp., 389.

TABLE V

Average Scores of the Normals in the Drill Series

Class	Drill Series 1 Av.	1 m. v.	2 Av.	2 m. v.	3 Av.	3 m. v.
I	49.059	3.64	56.046	3.37	54 276	3.81
II	42.737	2.28	47.852	5.29	47.618	5.34
III	40.423	2.64	43.839	5.02	44.325	5.50
Total	43.882	3.73	49.137	5.64	48.988	5.64
Girls	41.989	3.90	46.746	5.44	47.184	5.69
Boys	44.885	3.64	50.719	5.07	50.299	5.42

exactly contrary results were obtained with the normal children. In the present experiment the girls were mentally superior to the boys; in the earlier work the boys were mentally superior to the girls, yet in both cases the girls showed an improvement in *Drill Series 3* and the boys a decline. This difference, therefore, cannot be explained upon a basis of mental age; but appears to be a real sex difference, probably due to the girls sustaining their interest and effort throughout the experiment; whereas the boys probably lost interest toward the end of the work and made correspondingly less effort. This explanation is strengthened by the fact that the *mean variation* of the boys in both experiments is larger than that of the girls, showing at least that the girls' attitude was the more constant.

c. Effect of Practice in the Test Series: The effect of practice in the Test Series is clearly indicated in Table VI:

TABLE VI

Average Score and Mean Variation of the Abnormal in the Test Series

Class	Test Series A 2nd week Av.	A m. v.	B 7th week Av.	B m. v.	C 17th week Av.	C m. v.	D 17th week Av.	D m. v.	E 54th week Av.	E m. v.
I	43.200	2.27	46.306	2.06	48.432	2.03	49.728	2.46	49.325	2.08
II	38.686	1.97	41.262	2.18	42.969	2.57	42.692	3.34	41.834	3.01
III	32.366	2.53	33.378	3.43	36.401	4.85	37.010	4.47	34.675	4 07
Total	38.223	4.28	40.482	5.69	42.752	5.05	43.269	5.52	41.939	5.86
Girls	38.165	3.37	41.792	4.50	44.694	4.60	44.952	4.91	43.986	4.77
Boys	38.274	5.01	39.461	5.88	41.233	5.70	41.952	5.92	40.096	6.20

The subjects made a slow and gradual improvement throughout the entire experiment. This relation holds true whether the group is considered as a whole, by class, or by sex. The curve of improvement in all cases is relatively the same, showing a slow and continued rise. It differs, however, in this respect from the normal curve. For as was shown by Whipple[15] and Foster[16] in their work upon adults, ment in visual apprehension is very rapid. After a pronounced initial rise, which may be explained by reference to such determining factors as familiarity with the material; application of more efficient methods of work; the use of tricks of counting, grouping, and naming; and the direction and restriction of attention, there is no appreciable gain. Table VII, which is a reproduction of Table II of the earlier work,[18] gives the records of the normal subjects in the Test Series, and is comparable to Table VI.

TABLE VII

Average Score and Mean Variation of the Normal in the Test Series

Class	Test Series									
	A 2nd week		B 7th week		C 12th week		D 17th week		E 58th week	
	Av.	m. v.	Av.	m. v.	Av.	m. v.	Av.	m. v.	Av.	m. v.
I..........	38.662	3.64	48 256	3 97	48.108	3 32	48.243	2.99	52.400	1.66
II.........	32.266	2.86	43.166	3 64	43 400	2.68	42.650	3.75	45.700	3.77
III........	30.441	2 08	39.455	2.62	41.176	2.13	43.014	2.08	46.350	1 35
Total......	33.995	4.40	43.782	4 07	44.376	3.49	44.856	3 46	48 000	3.70
Girls.......	33.392	3.70	41.357	3.86	42.595	3.99	43 452	3.82	48.571	3.01
Boys.......	34.423	3 66	45.536	3.90	45.644	3.61	45 855	3.62	47.666	3.55

A comparison of Tables VI and VII reveals the fact that while the improvement of the feebleminded subjects is very different from that of the normal taken as a group, it nevertheless resembles very closely the improvement of the poorest third, *i.e.*, Class III.

The amount of improvement for both groups of subjects is shown for every Test Series in Table VIII.

[15]Op. Cit., 258 f.

[16]Op. Cit., 16 f.

and by us in our work upon school children[17] the normal improve-

[17]Op. Cit., 334 f.

[18]Pp., 389.

TABLE VIII

The Average Loss and Gain in the Number of Elements Apprehended in the Test Series

Class	Normal				Abnormal			
	Test Series				Test Series			
	B	C	D	E	B	C	D	E
I...........	+ 9.6	−0.1	+0.1	+4.2	+3.1	+2.1	+1.3	−0.4
II...........	+10.9	+0.3	−0.8	+3.1	+2.6	+1.7	−0.3	−0.8
III..........	+ 9.4	+1.7	+1.9	+3.3	+1.0	+3.1	+0.6	−2.4
Total........	+ 9.8	+0.6	+0.5	+3.2	+2.2	+2.3	+0.6	−1.3
Girls.........	+ 8.0	+1.2	+0.9	+5.1	+3.6	+2.9	+0.3	−1.0
Boys........	+11.1	+1.1	+0.2	+1.8	+1.2	+1.8	+0.7	−1.8

(Note: These data were obtained from Tables VI and VII.)

This Table clearly presents the points that we have just made, namely: throughout the drill, in *Test Series B, C,* and *D*, the feebleminded and the poorest third of the normal group continue to improve; whereas the superior and medium subjects of the normal group show a large initial increase and then a long plateau. We, therefore, conclude that the characteristic effect of practice in visual apprehension in the backward and feebleminded is a slow and gradual gain; and conversely, that a slow and gradual improvement in visual apprehension is indicative of backwardness.

d. Effect of Practice in the Different Types of Material: The effect of practice upon the different types of material is shown in Table IX.

TABLE IX

Average Number of Elements Apprehended in Each Test Series for Every Type of Material

Test Series	Type						
	I	II	III	IV	V	VI	VII
A.............	4.93	3.96	4.75	3.16	4.73	4.29	3.89
B.............	5.38	4.31	4.82	3.35	4.77	4.09	3.51
C.............	6.13	4.75	5.30	3.62	5.63	4.27	4.11
D.............	6.20	4.84	5.02	3.93	4.96	4.59	3.87
E.............	5.84	4.77	5.07	3.68	4.79	4.10	4.31

I.—Numerals; II.—Letters; III.—Numerals and letters; IV.—Words; V.—Numerals, letters and words; VI.—Geometrical figures; VII.—"Tit-tat-to" combinations.

A steady improvement can be noted in the average number of elements reported in Types I, II, IV, and VI—the pure types, viz., the numerals, letters, words, and the geometrical figures. There is also an initial improvement in the reports of Types III, V, and VII,—the mixed types, viz., numerals and letters; numerals, letters, and words; and the "tit-tat-to" combinations,—which persisted through *Test Series C*, but which was followed by an abrupt decline in *Series D*. Since no introspections were taken we are unable to account for this difference. The uniformity, however, of the improvement of the pure types, and the decline of the mixed types, indicates that this difference may be due to the complexity of the material and the diversity of the associations aroused.

The table also shows the average number of elements of each type which can be grasped in a 5–second exposure. Before the practice, the number ranges from 3.16 to 4.93; after the practice from 3.87 to 6.20. The largest gains are made in the "pure" types; for example: in Type I *Test Series D* 6.20 elements are apprehended, a gain of 1.27; inType III 5.02 elements are apprehended, a gain of only 0.27. The average gain in the pure types is 0.80, of the "mixed" types 0.12 elements.

e. Retention of Practice: Test Series E was submitted after an interval of 36 weeks in order to discover whether the effects of practice were permanent. Fortunately all but three of our subjects were able to serve in the re-test, and their results appear in detail in Table I. The average results and mean variations are shown for every class in Table VII.

Though all the classes show a loss there is surprisingly little difference between the results of this test and those of *Test Series D*. The net loss is slight, amounting, as shown in Table VIII, to only a little over 1 unit. This is principally due to the reports of the poorest third of the class and to the boys. The individual records, however, are very much superior to those made a year previously (before the drill), hence we may conclude that the effects of practice are relatively permanent.

2. Special Points:

a. The Effect of the Practice upon Mental Age: The children were re-examined at the conclusion of the Drill Series with the Yerkes-Bridges test in order to learn whether practice in visual apprehension had effected their mental status. The results of this re-test appear in detail in the fourth and fifth columns of Table I.

An inspection of these data reveals the fact that the subjects, with but two exceptions, make a higher record in the second than in the first examination. Don, in Class I, shows a loss of 0.42 years; and Freddie, in Class III, shows no change. The amount of improvement is more clearly shown in the gross class and group averages which appear below:

TABLE X

Results of Mental Examinations by the Point Scale of the Practiced and Unpracticed Groups

Class		Practiced Group				Unpracticed Group			
		First test		Second test		First test		Second test	
		Before practice		After practice					
		Mental age	C. I. A.	Mental age	C. I. A.	Mental age	C. I. A.	Mental age	C. I. A.
I	Av.	9.68	0.730	10.39	0.790				
	m. v.	0.76	0.069	0.92	0.064				
II	Av.	8.69	0.610	9.63	0.720				
	m. v.	0.40	0.065	0.90	0.067				
III	Av.	8.17	0.552	8.64	0.608				
	m. v.	0.33	0.045	0.49	0.065				
Total	Av.	8.86	0.637	9.57	0.710	9.14	0.644	9.28	0.650
	m. v.	0.67	0.085	0.99	0.095	0.96	0.099	0.82	0.086
Girls	Av.	9.21	0.652	10.24	0.741	9.08	0.620	9.24	0.650
	m. v.	0.82	0.098	0.093	0.090	1.23	0.126	0.95	0.090
Boys	Av.	8.59	0.626	9.05	0.684	9.19	0.661	9.32	0.668
	m. v.	0.58	0.087	0.67	0.095	0.71	0.074	0.69	0.085

C. I. A., Coefficient of intellectual ability.

This table shows that the average gain is almost a year (0.71). It is most noticeable among those having the highest coefficient of intellectual ability. Classes I and II, and the girls have the highest coefficients, and they also show the greatest gain.

In order to ascertain whether this gain was due to an actual mental improvement, and not to the mere repetition of the tests, a group of 11 children, six girls and five boys, was used as a control. These subjects were tested and re-tested at the same times as were those of the main experiment. The results of their examinations appear

also in Table X, under the caption *Unpracticed Group*. These children show on an average a very slight gain in the re-test; a gain—0.14 of a year—which we should expect solely as a consequence of the increase in chronological age.[19] Since the control, the unpracticed subjects, show no effects of repetition, the mental improvement of the practiced observers can not be explained upon a basis of familiarity with the mental tests. Whether this improvement is due to the training in our drill in visual apprehension, and, if so, whether the transfer is specific or general, we can not here say. The improvement, as we have remarked above, is very uniform, and is so large that it can not be neglected. These results confirm those of our earlier experiment,[20] and strongly indicate a transfer of training.

b. Relation to Mental Age: The classification of the subjects into the good, medium, and poor,—Classes I, II, and III respectively,—agrees very closely with a similar division of mental ages shown by an inspection of the data in Tables IV and X. Table IV shows the average standing of each class in visual apprehension; Table X shows the average mental age in each class in the two Yerkes-Bridges examinations. A comparison of these tables reveals that the good observers, Class I, have the highest mental age; and the poor observers, Class III, the lowest mental age. Likewise, as regards sex, the girls make the higher score in our work and they test the older mentally.

These results would seem to indicate a closer correlation between ability in visual apprehension and mental age,—and such a correlation is found to exist. Two positive coefficients show this relation: the first, between the subjects' standing in the *Preliminary Practice Series* and the first mental test; the second, between the subjects' standing in the final series of the training, *Test Series D*, and the second mental test. The coefficients and probable errors, as obtained by Spearman's method of rank differences, are:

	Correlation	P. E.
1. Standing in visual apprehension and mental age, before practice....................	0.70	0.056
2. Standing in visual apprehension and mental age, after practice.....................	0.63	0.094

[19]After an interval of 17 weeks (0.33 years) the group showed an improvement of only 0.14 year, which corroborates Yerkes's findings, "that the Point Scale appears to stand fairly satisfactorily this kind of test of reliability." (*A Point Scale for Measuring Mental Ability*, 1915, p. 102).

[20]Pp. 399–402.

c. The Effect of Practice upon Capacity to Observe: In order to discover whether the training was specific, whether our subjects had really become noticeably better observers, and whether there was a transfer in training, an *Aussage* test was given. As a check or control it was also given to the unpracticed group mentioned above.

These tests were conducted by D. G. Paterson, in June, 1916, immediately after the conclusion of the main experiment. He makes the following report:

"All the children, with the single exception of Walter R. of the practiced group, Class II, who was ill during this work, were individually subjected for 30 seconds to the Binet 'card of objects' test.[21] The reports of the two groups were compared first, with respect to the number of objects correctly recalled; and secondly, with respect to the number of answers correctly returned to an 'interrogatory' of 50 set questions.[22] The results show that the practiced group was not superior to the unpracticed in the number of objects correctly recalled; but was superior in the number of questions correctly answered. The data are given below in Table XI."

TABLE XI

Group Averages and Mean Variations in the Aussage Test

Groups	Classes	Objects Correctly Recalled		Questions Correctly Answered	
		Av.	m. v.	Av.	m. v.
Practiced	I.	4.30	1.14	29.0	7.3
	II.	4.17	1.33	26.9	3.6
	III.	4.15	1.20	26.9	6.2
	Av.	4.20	0.38	27.4	5.8
	Girls.	4.58	1.35	27.4	6.0
	Boys.	3.90	1.00	27.4	5.9
Unpracticed	Girls.	3.83	1.05	25.5	5.5
	Boys.	4.60	0.48	26.5	5.1
	Av.	4.18	1.06	25.9	5.3

[21] *Cf.* WHIPPLE, G. M., *Manual of Mental and Physical Tests*, 1st Ed., 1910. Test, 32, A, pp. 297-301.

[22] *Cf.* WHIPPLE, G. M., *Ibid.*, 298-299.

The superiority of the practiced group in the number of objects correctly reported is very slight, and of no particular significance being probably a chance variation, for the difference is only 0.02, with a probable correctness of only 0.55.[23]

The difference in the number of questions correctly answered, was, however, more marked. The average score of the practiced group was 27.4; that of the unpracticed group 25.9. The difference, 1.5, has a probable correctness of 0.77, and is therefore not likely of chance occurrence. Still, we hesitate to conclude that this is due to training in our sort of work, for a comparison of Tables X and XI reveals that the capacity to observe is, as shown above in the preceding section, closely correlated with mental age: the practiced group is mentally older than the unpracticed group; Class I gives the best reports and is the oldest mentally; Class III gives the poorest reports and is the youngest mentally; the girls, in the practiced group, give better reports and are older mentally than the boys; and the boys, in the unpracticed group, give better reports and are older mentally than the girls.

Were it not for the results of Class III we should be satisfied to conclude that the difference in the capacity to observe is due to a difference in mental age, and that there is no evidence of a transfer of training. Class III is mentally over half a year younger than the control group and yet gives better reports in the *Aussage* tests. The differences and the probable correctnesses are small, and may be regarded as chance variations. Nevertheless, the results point to a transfer of training, for we still have to explain why the records of this group equal those of the control group, when according to mental age we should expect them to be inferior. It may be, and we offer this merely as a tentative explanation, that the reports in the *Aussage* test are conditioned both by practice and by mental age; and that the effects of the former are obscured by the latter in all the groups except Class III.

d. Comparison with Adults: In the comparison of the range of visual apprehension of practiced feebleminded children with adults, one of the most difficult sections of *Drill Series 3*, one in which there were 100 elements, was given to an elementary class in psychology at Ohio State University. There were 94 students in the class, composed mostly of freshmen, 28 women, and 66 men. The data thus obtained were graded according to the method described above,

[23]Calculated according to the formula of E. G. Boring; *On the Computation of the Probable Correctness of Differences, Am. J. of Psy.*, XXVIII, 1917, pp. 454–459.

and the students were then divided upon a basis of these grades into three groups, the good, the medium, and the poor, Classes I. II, and III, respectively. The average record of each of these classes, the total average, and the average according to sex, appears in comparison with similar records of the children in Table XII.

TABLE XII

Comparison of the Range of Visual Apprehension of the Feebleminded Children and the Adults

Class	Feebleminded Subjects		University Students	
	Av.	m. v.	Av.	m. v.
I.............	55.8	4.0	67.3	2.7
II............	45.3	1.9	59.1	1.5
III...........	35.9	3.9	50.7	2.9
Total.........	45.47	7.21	58.84	4.64
Girls..........	48.7	8.2	56.1	5.6
Boys.........	43.0	6.3	60.0	5.6

These results show that the adults are on the average better observers than feebleminded children, even though the children have had four months of special training. It is also worthy of note that the feebleminded children of Class I excel the adults of Class III; that is to say, the upper third of a group of feebleminded children give better reports than the lower third of a class of University students.

e. Dependence upon Chronological Age and Sex: Though the general results of practice were identical for the different ages and sex, the range of visual apprehension varies with chronological age and with sex.

The correspondence with chronological age, however, was not absolute. This is shown by the results of Table XII, and by a comparison of the results of Table III and IV. The adults, whose average ages were 19.3, give better records than the feebleminded subjects whose ages averaged 14.6; the girls, whose chronological ages averaged 15.59 give better reports than the boys whose chronological ages averaged 13.03; and Class II, the group of medium observers, was chronologically older than Class III, the group of poor observers. Here, however, the correlation between chronolog-

ical age and performance in the tests of visual apprehension ends; for Class I, composed of the best group of observers, is chronologically the youngest,—yet withal, it is the oldest mentally. Since it is the oldest mentally, and since the other groups which we have mentioned above as excelling are likewise the oldest mentally, we must conclude that the range of visual apprehension is correlated with chronological age only in so far as it is a condition of mental age.

Under the conditions of this experiment the girls were better observers than the boys. The results leading to this statement are quite uniform: the girls records are superior to those of the boys in the *Drill Series* (Table IV), in the *Test Series* (Table VI), in the amount of the practice gain (Table VIII), in the retention of the effects of practice (Tables VI and VIII), and in the degree of variability as shown by the mean variations in Tables IV and VI. This superiority, however, we do not believe to be on account of a sex difference, but in spite of one. The influence of sex, we think, is observed by other factors. The girls are chronologically 2.56 years, and mentally 1.19 years older than the boys, and both of these factors, as we have shown above, condition the range of visual apprehension; and we believe, therefore, that they are responsible in the present instance for the girls' superiority. In our earlier work[22] exactly contrary sex results were obtained; the boys in the normal group were superior to the girls, and the men in the adult experiments (those reported above and those reported in our earlier work),[23] were superior to the women. Class I of the present adult experiment is composed of 7 women and 24 men, Class II of 6 women and 25 men and Class III of 15 women and 17 men. When the members of this group are arranged in order of rank, there is not a woman among the first 11. So with the feebleminded. The male superiority is shown by the fact that when subjects of the same average mental age are chosen and their records compared, those of the boys are slightly greater than those of the girls. Therefore we conclude, that men and boys of the same mental status are slightly superior in visual apprehension to women and girls.

f. Individual Differences: The individual differences in the capacity for visual apprehension are large, as is clearly indicated by Table I. Some of our subjects are very efficient, even excelling adults; others very poor. The range of the reports extends, in the *Preliminary Practice Series*, from 57.5 to 19.5; in the *Drill Series*,

[22]Pp. 395 f. [23]Pp. 394 f.

from 69.5 to 19.0; and in the *Test Series*, from 59.3 to 18.5. These differences, however, offer no particular problem, for they may be satisfactorially explained upon a basis of mental age.

g. Variability: The degree of variability in visual apprehension as expressed by the m. v. is shown for the three classes, for the total, and for the girls and boys, in Tables IV, V, VI, VII, and XII. Table XII shows that the adults have a smaller m. v. than the children; Tables IV and VI show that of the feebleminded group, the girls have a smaller m. v. than the boys, and that Class I has on an average the smallest m. v. and Class III the largest; and Tables V and VII show that of the normal group the boys have a smaller m. v. than the girls. Since the adults are mentally older than the children, the feebleminded girls than the feebleminded boys, Class I, than either Class II or III, Class II than Class III, and the normal boys, than the normal girls,[24] we conclude that the degree of variability is conditioned by and varies inversely as the mental age.

A comparison of Tables IV and V, and Tables VI and VII, however, reveals that the feebleminded children, who are mentally older than the normal children (Table III) have of the two groups the larger m. v. This seems to be in direct contradiction to the conclusion drawn above. In reality, however, it is not; it is rather the result of a second condition which in this instance cuts across and obscures the condition of mental age. The effect of this second condition is also shown in Table X where it appears that Class III, which had the largest m. v., had also the smallest coefficient of intellectual ability; and Class I, which had the smallest m. v., had the largest coefficient. Since the abnormal have a larger m. v. than the normal, and the subjects with a low coefficient than those with a high coefficient, we conclude that the extent of variability is conditioned by and varies directly as the degree of abnormality.

The conditions, then, that effect variability are mental age and abnormality. These two conditions may augment the variability, as in the case of the several classes; or the effects of the former may be obscured by those of the latter, as in the case of the normal and abnormal subjects.

[24]As we have said before, the normal children unfortunately were not given a mental test. We came to the conclusion that the boys were mentally older than the girls because their records in the practice and tests of visual apprehension surpassed those of the girls. Since we found, as shown above, that performance in visual apprehension is directly correlated with mental age, it follows that the boys were in fact, mentally older than the girls.

h. Weekly Rhythm: The average of the reports in the *Test Series* was computed for each day of the week. Little evidence was obtained for weekly rhythm. It was found that the average range of visual apprehension in a 5-second exposure, was 4.11 on Monday, 4.10 on Tuesday, 4.16 on Wednesday, 3.99 on Thursday, and 4.10 on Friday. There is a slight drop in the subjects' capacity on Thursday which we are unable to explain, but the significance of the difference is so small that it is due probably to a chance variation.

III. SUMMARY

The general results of this study may be summarized as follows:

(1) The effect of practice in visual apprehension in the backward and feebleminded children is characterized by a slow and gradual improvement; conversely such an improvement in visual apprehension is indicative of backwardness.

(2) The range of visual apprehension varies with the complexity of the material and the diversity of the associations aroused.

(3) The effect of the practice is relatively permanent.

(4) The subjects' improvement in mental age, and their superiority in their capacity of general observation is evidence for a restricted belief in formal discipline.

(5) There is a direct correlation between the range of visual apprehension and mental age.

(6) Adults, university students, tested without special practice, are superior to trained feebleminded subjects, though the lower third of the adults are then inferior to the upper third of the feebleminded.

(7) The range of visual apprehension is correlated with chronological age, only in so far as chronological age is a condition of mental age.

(8) When subjects of the same mental age, or status, are compared, men and boys are slightly superior to women and girls.

(9) Individual differences are marked, but they are closely correlated with mental age.

(10) The extent of variability is conditioned by mental age and degree of abnormality, varying inversely with the former and directy with the latter.

THE ABILITIES OF PUPILS IN DETROIT PREVOCATIONAL CLASSES

SAMUEL RENSHAW

Western State Normal College, Kalamazoo, Michigan

The prevocational classes in the City of Detroit were formed for the purpose of providing instruction and training for pupils who have been "unable to profit by the instruction in the ordinary public elementary schools" and who have spent some time in a special class or ungraded room and have now reached the adolescent years. The immanence of their reaching the upper age limit of the compulsory education law forcibly presents new problems for those whose duty it is to do all that can be done, in a scholastic way, for these atypical children. The seven prevocational classes were instituted to add, in a last Herculean effort, whatever possible to make these pupils more nearly conformable socially and to help them to attain the ultimate limits of their capacities. There are five such classes for boys, and two for girls.

The purpose of this investigation was to collect information which might throw some light on such questions as "How nearly, with present curricula and methodologies, do these classes fulfill their mission?" "With pupils of this intellectual calibre is it profitable to continue the present traditional admixture of manual training, industrial arts and academic subjects (geography, reading, language, arithmetic, spelling, etc.)?" "Are these classes affording a type of instruction that, with pupils such as these, will function in a truly prevocational way?" "Are the efforts expended here producing results in turning these children out at sixteen better fitted to become self-sustaining citizens?" Or, "Are they merely marking time, with these subjects serving as busy-work, until they pass beyond the school age and the burden of responsibility for the pupils is shifted to other shoulders?" The only satisfactory means of approaching these questions and a legion of others that obtrude upon the minds of those whose business it is to administer such

schools is to determine the specific abilities and disabilities of the subjects who are receiving this course of training and instruction. Provision for the studies reported here were made by Mr. Fred. L. Keeler, Superintendent of Public Instruction, Lansing, Michigan, and Mr. Frank Cody, Assistant Superintendent of Schools, Detroit.

The Subjects

The subjects comprised about 300 pupils in the five boys' classes, Nichols, Cary, Russell, Goldberg and Newberry, and the two girls' classes, Lincoln and Craft. This number is probably too low and does not include all pupils of this age and status in the city. The total school population is 97,740. Grade-age reports from the department of statistics show that 5.3% of these 97,740 pupils are 3 years or more retarded. At present there are 35 special classes in all, enrolling 738 different pupils, averaging about 22 per class. Instead, assuming that 2% of the total school population are of this level, it would mean there should be not less than 1954 pupils in these special classes. It would take 74 such classes of 25 pupils each to accommodate this number, which would mean the establishment of 39 additional classes, of which at least one third would have to be prevocational classes. *Cleveland, Ohio, of slightly smaller population, had in 1915 one hundred and forty special classes and enrolled 2,559 pupils.

The chronological ages of the pupils ranged from twelve and one-half to seventeen, with the median age at about fourteen and one-half. The number present in each of our studies was not constant, for various reasons. The number varied from about 264 to 308.

The Tests

It was decided to ascertain the level of general intelligence of these pupils, their linguistic ability, their ability to read silently, and to deal with numbers. We secured ratings of each with the Stanford Revision of the Binet scale, Trabue Completion Test Language Scales B and D, (D, of equal weight and difficulty, being used as a check) the Kansas Silent Reading Test (Kelly's Group 1) and Starch's Arithmetical Scale A (Reasoning). We assumed that these quantifications would afford a reasonably justifiable description of intellectual and pedagogical equipments. The latter three

*David Mitchell. *Schools and Classes for Exceptional Children.* Cleveland Survey Report.

tests were given by the regular class teachers after instruction and training. All scoring and tabulating of results was done by the writer. The Stanford Revision was given to each pupil during the course of the school year by Misses Alice B. Metzner and Anna Engel, the mental examiners for the board of education of the City of Detroit.

Trabue's Completion Test Language Scales were selected as measures of the ability to use and control the mother tongue because of their facility in scoring and their high correlation with general intelligence. *Thorndike, Stenquist and Trabue found this to be .875 with the Goddard 1911 Revision with about 150 dependent subjects in ages 10 to 14. In our cases a random sampling showed a Spearman "r" to be .92 with the Stanford Revision ratings. A glance at Table X shows how closely the percentiles approximate in each age year. Scale B was followed about 3 weeks later with scale D as a check and to determine the nature and extent of any practice effect. A comparison of the distribution in Table IV shows this to be a practically negligible factor in the case of these pupils—a fact not without its significance.

Since the ability to read silently with speed and comprehension is the one *sine qua non* in most school subjects we decided to use Kelly's Kansas Silent Reading Tests—Series 1—as the measure of this ability. The question as to whether these tests are only examples of problem solving in arithmetic, hard-directions tests, or whether they elicit performances from which abilities are inferred is irrelevant to our problem. Subnormals give quite a distinctive type of performance with this test and some interesting side-lights on the mental make-up of these pupils are revealed, as our results show. The tests were new to the pupils and were given in the regular reading exercise period. The same is also true of the measures of language and arithmetic.

Because of its simplicity in scoring and the allowance of plenty of time (30 to 45 minutes) and because it involves not only the fundamental operations but also the solution of many practical examples, we selected Starch's Arithmetical Scale A (Reasoning) for the work with numbers. His June standards were employed in evaluating results. It might be added that there was very little uniformity

*Psychological Clinic, Vol. X, No. 5.

in the amount of time, methods of presentation or stress upon arithmetic in the seven classes. In some of the schools twice as much time was devoted to certain subjects as in others.

Results

Intelligence Status of the 308 Pupils

The following table shows the ratings of the 308 pupils in these classes. The number rating

I. Q.	Diagnosed as—	% of total cases
.70 – .90	Backward	38.94
.50 – .70	Feebleminded	54.21
.20 – .50	Idiot-Imbecile	3.56

But 6.81 per cent. of these cases rate between I. Q. .80 and .92. Practically the only ones from whom anything like useful independent citizenship can be expected must be drawn from this small per cent. of the total group. As will be shown later but about 12 per cent. make scores equal or in excess of the level of the normal 5th grade child in the language completion tests, in arithmetic but about 2 per cent., and in silent reading this number is less than 1 per cent. of the whole number. Sixty-three per cent. of all these pupils have an I. Q. below .70.

Tables II and III show the distribution of the chronological and mental ages. Eighty-nine per cent. of the cases lie in mental ages VIII, IX and X, while the median chronological age is a few months in excess of fourteen years. Tables I, II, and III show the numerical distributions by classes. The 5 boys' classes enroll about

TABLE I

Numerical Distribution of Intelligence Quotients

	.25 to .40	.40 to .49	.50 to .59	.60 to .69	.70 to .75	.76 to .79	.80 to .92	Total
Nichols	0	0	2	15	12	7	6	42
Craft	0	1	7	23	11	3	2	47
Goldberg	0	1	5	7	18	2	4	37
Russell	1	1	10	22	11	5	0	50
Lincoln	1	4	11	17	7	0	1	41
Cary	0	0	6	26	7	4	7	50
Newberry	0	2	7	19	12	0	1	41
Total	2	9	48	129	78	21	21	308

TABLE II

Chronological Ages

	XII	XIII	XIV	XV	XVI	XVII
Nichols....	0	10	19	12	4	0
Goldberg...	1	4	10	10	2	1
Newberry...	0	2	19	8	3	0
Craft......	0	6	14	18	5	0
Russell.....	2	6	13	16	1	0
Cary.......	0	8	13	13	1	0
Lincoln.....	0	5	14	14	3	1
Total..	3	41	102	91	19	2

TABLE III

Numerical Distribution of Mental-Ages

M. A........	5	6	7	8	9	10	11	12	13	14	15	No. Cases
Newberry.......	1	1	6	11	17	5	0	0	0	0	0	41
Cary............	0	0	3	10	18	11	5	0	1	1	1	50
Lincoln.........	0	2	6	14	12	4	0	0	0	0	0	38
Russell..........	1	0	4	15	16	7	2	0	0	0	0	45
Goldberg........	0	0	4	6	11	11	5	0	0	0	0	37
Craft...........	0	1	3	12	17	7	3	0	1	0	0	44
Nichols.........	0	1	3	12	19	12	4	1	0	0	0	52
Total.......	2	5	29	80	110	57	19	1	2	1	1	307

Median Mental Age.... 9.3 years
Median Chron. Age.....14.5 years (approx.)

180, and the 2 girls' classes about 90. Whether this supports the view that girls are more "pure line"—that they deviate less than boys—cannot be said. It may be due to the mode of selection, or differences of criteria, or to other factors.

TABLE IV

Distribution of Scores—Trabue Completion Language Scale B

Scores...	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Median
Russell......	4	0	3	1	3	4	5	1	3	0	2	4	2	4	0	0	6.60
Craft.......	0	0	4	0	2	2	16	2	9	5	2	0	1	0	0	0	6.84
Nichols....	0	0	0	0	0	3	3	1	6	8	13	2	2	4	3	0	10.12
Newberry...	1	0	1	0	10	1	7	4	6	1	2	2	0	0	0	0	6.64
Goldberg....	2	0	1	0	5	0	7	2	4	3	3	1	0	0	0	0	6.85
Cary........	2	0	2	1	3	1	0	7	8	2	5	3	0	1	0	1	8.33
Lincoln.....	6	0	1	1	9	0	10	2	3	3	0	0	2	0	0	0	6.15
Total...	15	0	12	3	32	11	48	19	39	22	27	12	7	9	3	1	7.21
Scale D Total..	21	1	22	8	32	7	28	24	37	11	24	16	8	4	1	0	7.25

TABLE V

Quartile, Median and Upper and Lower Quartile Scores, Trabue Scale B

	Lin.	Russ.	New.	Craft.	Gold'g	Cary	Ni.
75 percentile....	7.37	11.25	8.37	8.69	8.99	10.20	10.98
Median........	6.15	6.60	6.65	6.84	6.85	8.32	10.12
25 percentile....	4.13	4.33	4.60	6.10	4.80	5.99	8.70
Quartile........	1.62	3.46	1.88	1.29	2.09	2.10	1.14

TABLE VI

Languages Spoken in the Homes of Pupils in Six of the Seven Schools

	New.	Cary.	Gold'g	Lin.	Russ.	Craft.	Total
English........	18	34	24	9	14	20	119
English and Foreign.......	5	9	1	9	10	9	43
Polish..........	7	3	3	2	7	10	32
German........	7	0	0	2	0	2	11
Russian........	2	0	0	0	0	0	2
Bohemian......	1	0	0	0	0	1	2
Hungarian......	0	2	0	1	0	0	3
Italian.........	0	0	1	6	5	0	12
Jewish.........	0	0	3	6	12	0	21
Russian-Jewish..	0	0	0	10	0	0	10
Austrian........	0	0	0	1	0	0	1
Lithuanian.....	0	0	2	2	0	0	4
Syrian.........	0	0	0	1	1	0	2
French..	0	2	0	0	0	0	2
Norwegian......	0	0	1	0	0	0	1
Finnish.........	0	0	1	0	0	0	1
Deaf-Mute.....	0	0	1	0	0	0	1
Total	40	50	37	49	49	42	267

Tables IV, V and VI show the results of the Trabue scales. Here the Nichols pupils stand distinctly superior in abilities to the other groups. Nichols is a "favored" school in a better residential section and this may account in part for the superiority. Table VI shows the languages spoken in the homes of these pupils. Records of the Nichols group are not included as they were not available. The data were secured by the teachers in visits to the homes. In 56 per cent. of the homes some tongue other than English is the vernacular. There is a large foreign population in Detroit. What influence this has on the performances of these pupils in educational measurements it is difficult to say. Most of the pupils, however, have spent at least 5 years in American schools. Seventy-five per cent. of all the pupils rate below the normal III grade level in these tests. The lower quartile group rate even below the median of the normal second grade. A few cases show abilities as high as those found in high school classes. The medians of the two girls' classes stand considerably below the boys' medians.

In arithmetical ability (see Tables VIII and IX) fully 80 per cent. of all cases lie below the normal V grade level. One hundred twenty-four of the 270 pupils taking this test rated below or equal to the normal III grade standard. The best group, Nichols, are about twice as capable in arithmetic as the poorest group, Lincoln. Miss Merrill[1] concludes from her testing Minnesota defectives with the Courtis B tests "The fundamental operations can only be taught with profit in the case of high grade morons and in a few instances with mid-grades."

TABLE VII

Numerical Distribution of Scores—Kansas Silent Reading, Test 1

Scores.	0	1.2 to 4.0	4.2 to 5.3	5.8 to 9.3	9.9 to 13.2	15.7
Nichols....	15	13	4	5	0	0
Craft.....	10	6	3	6	1	1
Lincoln.. ..	13	16	3	3	1	0
Newberry...	18	20	3	2	0	0
Russell.. ..	14	13	5	4	1	0
Cary.......	15	13	6	3	3	0
Goldberg..	8	12	6	1	1	0
Total..	93	93	30	24	7	1

III Grade Level (5.3)	IV Grade Level (9.3)	V Grade Level (13.2)

[1] MERRILL, MAUDE A., *Abilities of Special Class Children in the 3 R's*, Ped. Sem., XXV, 88-96.

In ability to read silently, as measured by this scale, of these pupils,

86% (216) fall below the III Grade Level.
14% (35) stand above III Grade Level.

Of this 14%—

5 pupils = III Grade, or 1.96% of all cases.
22 pupils = IV Grade, or 8.82% of all cases.
7 pupils = V Grade, or 2.80% of all cases.
1 pupil = VI Grade, or 0.42% of all cases.

TABLE VIII

Distribution of Scores—Starch's Arithmetical Scale A

Step	0	1	4	6	7	8	9	10	11	12	13	14	15
Value	0	.4	3.8	5.9	6.7	7.7	9.2	10.3	11.3	11.7	12.9	14.2	15.1
Nichols	0	1	2	5	5	10	10	4	1	0	0	0	0
Goldberg	4	2	3	8	5	6	3	1	1	0	0	0	0
Russell	4	6	11	6	10	2	4	0	0	0	0	0	0
Newberry	3	3	4	11	5	8	5	1	0	1	0	0	0
Craft	1	5	13	8	7	5	0	1	0	0	0	0	0
Cary	3	3	11	8	7	2	1	0	0	0	0	0	0
Lincoln	5	12	8	7	4	3	0	1	0	0	0	0	0
Total	20	32	52	53	43	36	23	8	2	1	0	0	0 —270

Standard June Scores

Grade	3	4	5	6	7	8
Score	4.5	6 2	7.8	9.4	11.0	12.6

TABLE IX

Median Scores in Arithmetical Scale

Nichols	8.60
Goldberg	6.65
Russell	6.62
Newberry	6 56
Craft	6.00
Cary	5.95
Lincoln	4.55

TABLE X

80% of the Total 270 Pupils are Distributed in Mental Ages VIII, IX and X. The Following Table Presents Percentages of Pupils Who in the Tests Actually Attained Age and Grade Scores Equivocal to Expectations from their Mental Ages

M. A.......	VIII	IX	X	Total %	% Below III Grade	% Above V Grade
Grade.......	III	IV	V			
Arithmetic..........	18.0	29.0	11.0	68	30	2
Silent Reading......	1.9	8.8	2.8	13.5	86	0.5
Language...........	25.5	24.0	10.3	59.8	26.9	13.3
Total M. A. percents. at this age......	26.6	36.6	19.0	= 82.2% of all cases.		

Table VII shows the performances of these pupils with the Kansas Silent Reading Tests. Lack of ability to comprehend word meanings and to follow simple directions seems to be one of the principal defects of these pupils. Eighty-six per cent. lie below the normal III grade level in this respect. But 24 attain the IV grade standard and only 7 reach the V grade standard. Weidensall,[1] in a similar way, says of her Bedford Hills subjects "a simple idea, a direction, a perception, does not penetrate and lead out into successful execution for any save the better third of our subjects . . . . " in her modification of the Woolley Cincinnati working girl vocational tests.[2] Table X represents the age and grade levels of attainment at the various mental ages. In a sort of stereotyped way these pupils are able to do the mechanical operations in arithmetic, that is to say they may be trained to multiply, add, subtract and divide. They are hardly educable with numbers. Where processes of abstraction and generalization are indicated it is a well known fact that these pupils are almost wholly unable to do anything worthy of mention with such exercises. For instance, at the Vineland Training school John is asked "Suppose you fed 6 ears of corn to your horses"—at which John interrupts "But we don't feed them 6 ears." "But just supposing you fed them 6 ears." "We never give 'em 6 ears, though." He

[1]WEIDENSALL, JEAN, *The Mentality of the Criminal Woman*, Educational Psychology Monographs, No. 14.

[2]WOOLLEY AND FISCHER, *Mental and Physical Measurements of Working Children*, Psy. Rev. Mon. XVIII. I, Whole No. 77.

was then asked "Suppose you took twelve ears of corn and fed them ten, how many would you bring back?" The answer was "Two." "How many are 12 less 10?" The reply was "Oh, I don't know that many." With pupils in whom there is such lack of insight, such marked deficiency in abilities of abstraction and generalization, in the perception of symbolized relations, and in comprehension of word meanings beyond the simplest nominal and verbal designations, how are we to expect them to be other than parties to a very uneconomical educational scheme at $125 per capita per annum?

What becomes of these pupils one, three, ten years after they leave school? Unfortunately there have been few investigations to follow them up. Pintner and Toops[1] in their studies of the abilities of unemployed men in the Free Employment agencies of Columbus and Dayton, Ohio, found 82 per cent. of the 94 men seeking employment as non-skilled laborers rated backward, border-cases or definitely feeble-minded. A comparison group in the two workhouses of these cities showed approximately the same distribution. Recent studies have shown that a great many recidivists in Sing Sing never had a trade—a steady job. It would seem then that the public schools could do a great deal if, rather than try to do the impossible with "academic" subjects, each of these pupils could be taught or trained to do some one thing with creditable skill, provided there were no indigenous limitations to prevent. There are reasons to believe, however, that much of the manual and industrial teaching in these classes is open to the same sort of criticism as the "academic" instruction. Along with Judd, in the case of the normal secondary schools, this "unfortunate dualism" of *manual* and *academic* could profitably be dispensed with. The somewhat common notion that all defectives can be trained to acquire a high degree of skill in the manual and industrial arts is not borne out in fact. It is just as reasonable to assume that all medical students, if given sufficient practice, can acquire a masterly surgical technique. That, however, is not the case. There is need for further studies of this problem. Again, it cannot be said that all "literary" instruction in these classes should be done away with. Just what type and how much are questions to be settled by careful and painstaking experimentation.

[1]PINTNER AND TOOPS, *Mental Tests of the Unemployed Jr. Appl. Psy. I, No. 4*, pp. 325 ff. II, pp. 15-25.

Summary and Conclusions

The justifiability of teaching "academic" subjects to *all* adolescent defectives in prevocational classes is to be questioned. If done at all, it can probably be done successfully only in the case of those who stand above the 90 percentile of this group.

Somewhat less than one child in every four can be said to be receiving practical benefit in a direct way from the instruction in reading, geography, language, arithmetic, etc.

Pupils of essentially the same mental age and I. Q. rating vary widely in linguistic, reading and arithmetical abilities. Hence the grouping of pupils of the same mental ages is not a reliable means of insuring proper educational treatment.

With abilities so variable, the handling of 30 to 50 pupils in the same instructional group forestalls individual instruction and defeats one of the aims of the special class. Likewise it is bound to entail great economic loss.

Only a very small per cent. of these pupils are intellectually capable of becoming independently self-sustaining. It is improbable that these and similar classes are truly *prevocational.*

Of the 308 cases 68 per cent. can do work in arithmetic comparable to children of the same mental age; 59.8 per cent. in language, and 13.5 per cent. in silent reading.

ENGLISH AND MATHEMATICAL ABILITIES OF A GROUP OF COLLEGE STUDENTS

EDWARD CHACE TOLMAN
University of California
From the Psychological Laboratory of Northwestern University

The following note presents the results of an attempt to determine to what extent "mathematical ability" and "English ability" are separate and measureable traits, in a group of college students.*

An introductory class in psychology made up of sophomores, juniors, and seniors (men and women), were asked to fill out cards stating whether they thought they had more ability in English or in mathematics. The reports were in the form "Much Greater," "Greater," "Slightly Greater," in either English or mathematics, or "Same." No definitions of the two abilities were given. Yet all but some half dozen of a class of 70 students could distinguish between the two sufficiently to be able to judge in which one they were better.

A week later the following 8 group tests were given.

Opposites.

39 stimulus words taken from the standardized list of King and Gold[1] varying from easy to hard. Time-limit 3 min. A credit of 3, 2, or 1 given to individual words.

Arithmetic.

A series of 13 problems mostly in subtraction, but a few involving addition, multiplication and division. All operations to be done mentally. Time-limit 3 min. A score of 1 for each correct answer.

Completion.

21 sentences taken from Trabue's[2] scales, varying from easy to hard. Time-limit 6 ½ min. A score of 1 given to each correctly filled blank.

---

*Considerable evidence has already appeared suggesting the relative independence of these two abilities. Bronner (Augusta F. Bronner, "The Psychology of Special Abilities and Disabilities," Little Brown and Co., 1917) has presented striking instances among clinic children of special abilities and disabilities sometimes in the sphere of language and sometimes in that of number-work. Kelley (Truman L. Kelley, "Educational Guidance. An Experimental Study in the Analysis and Production of Ability of High School Pupils," Teachers' College Contributions to Education, No. 71, 1914.) has developed regression equations for prognosticating High School grades, which were different for English and mathematics.

[1]KING, I, AND GOLD, H. *Tentative Standardization of Certain Opposites Tests.* Journal of Educational Psychology, 1916. 7, 459–482.

[2]TRABUE, M. R. *Completion Test in Language Scales.* Teachers' College Contributions to Education, No. 77, 1916.

Algebra.

This test was originated by the writer. It is printed in full. Time-limit 3 min. Each correct answer given a score of 1.

For each of the following conditions state the relation of one letter to another as required in terms of equal to ($=$), greater than ($>$), or less than ($<$).

Example:

Given	State relation of	Answer
$a=b;\ b=c$	c to a	$c=a$

If you cannot do one problem pass on to the next. Work as rapidly as possible.

Given	State relation of	Answer
$a>b;\ a<c$	b to c	
$c=b;\ b>k$	c to k	
$x+y=z;\ x>1;\ y=1$	z to 2	
$y=0;\ x+y=z$	x to z	
$a-b-c-d=0$	c to a	
$a+1=x;\ x+1=y$	a to y	
$a=b;\ b=c;\ c>d$	d to a	
$a<b;\ b<c;\ d>c$	d to a	
$a<c;\ b>c;\ b>d$	d to a	
$a-b=1;\ b-c=d$	d to a	

Rhymes.

20 2-syllable words, alternately accented on the first and on the second syllables, as follows: precede, ample, delay, pencil, delight, weather, destroy, mixture, affair, wagon, assess, nourish, arrange, butter, defend, churlish, review, hardy, commit, kettle. A rhyming word to be found for each. Time-limit 1 ½ min. Each correct rhyme scored 1.

Generalization.

One of the Courtis Geometry Tests.[1] The common element to be picked out in each of 10 sets of 5 words.

Illustrations

Set A. Mental, method, matter, muscle, mother.
Answer: All the words begin with the letter m.

Set B. Washington, Lincoln, Taft, Wilson, Jefferson.
Answer: All the words are names of Presidents of the United States.

Time-limit 2⅓ min. Each correct answer scored 1.

Sentences.

8 short sentences,* the words arranged in a meaningless order to be re-arranged to make sense. Time-limit 2¼ min. A score of 1 for each sentence correctly written.

*Three of the eight sentences were those given in the Binet scale.

Problems.

12 practical problems involving arithmetic to be solved mentally. Time-limit 4 min. A score of 1 for each correct solution.

[1]Courtis Standard Research Tests. Series G. Form 1. Test No. 4. S. A. Courtis, Detroit Mich.

The opposites, completion, rhymes, and sentences tests were supposed to be "English," the arithmetic, algebra, generalizations, and problems tests, "Mathematics."

The results were not computed for the half-dozen or so subjects who had reported "Same" ability for mathematics and English or for those who followed directions imperfectly. This left as the total number who took the tests and for whom the results were computed, 43. Of these 27 (2 men and 25 women) had judged themselves better in English and 16 (7 men and 9 women) better in mathematics.

The ogives, Figs. 1 to 8,[2] for each of the 8 tests and for the two groups, those who thought themselves better in English and those who thought themselves better in mathematics, have been drawn separately. Of the 4 supposedly English tests, the rhymes and sentences tests were apparently the most successful, the completion test somewhat less so, and the opposites test not at all. Of the 4 supposedly mathematics tests, the arithmetic and algebra tests were decidedly successful, the problems test less so, and the generalization test not at all.

The results for the 6 most successful tests will be considered further. Their intercorrelations are given in Table I.

TABLE I

	Completion	Sentences	Rhymes	Problems	Arithmetic	Algebra
Completion						
Sentences	.04					
Rhymes	.39	.13				
Problems	.36	.12	.03			
Arithmetic	.27	.16	.09	.34		
Algebra	.35	.15	.23	.33	.48	

The probable errors range from ±.10 for the correlation of .03, to ±.08 for that of .48.

Examining the table, the following points appear. The three mathematics tests correlate fairly well with one another, particularly the arithmetic and algebra tests (.48). In the case of the intercorrelation of the English tests, on the other hand, the only

[2]In all the charts the figures on the base line indicate the scores received in the test. The numbers in the vertical column to the left indicate the percentage of the group securing that particular grade or worse.

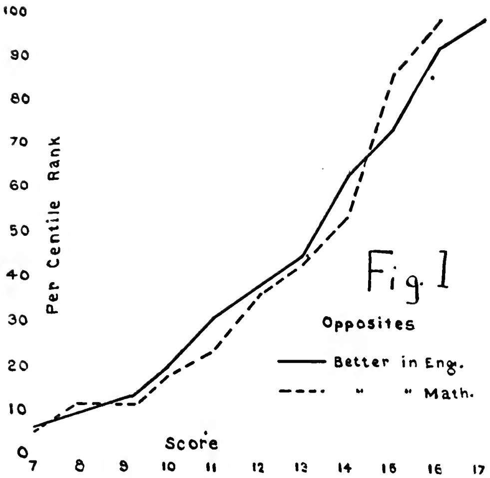


Fig. 1
Opposites
Better in Eng.
" " Math.
Per Centile Rank
Score
100 90 80 70 60 50 40 30 20 10 0
7 8 9 10 11 12 13 14 15 16 17

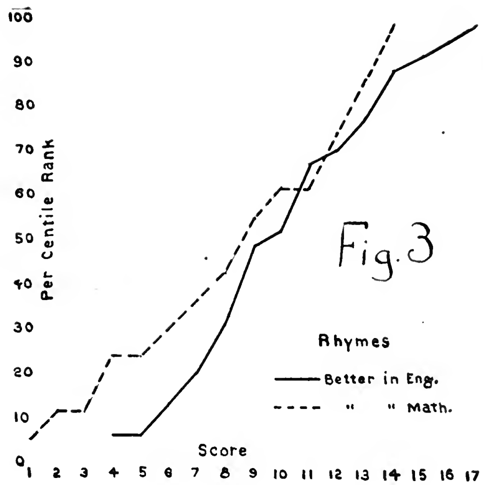


Fig. 3
Rhymes
Better in Eng.
" " Math.
Per Centile Rank
Score
100 90 80 70 60 50 40 30 20 10 0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

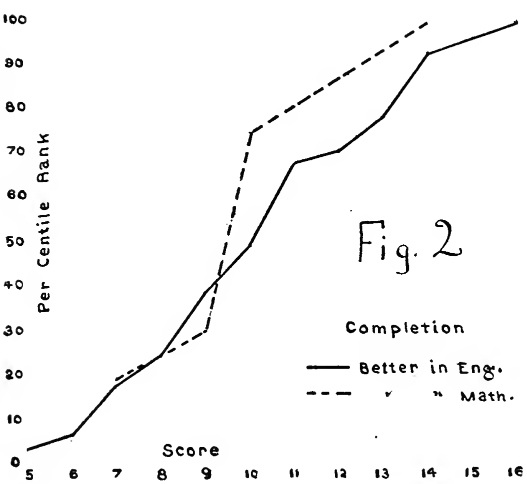


Fig. 2
Completion
Better in Eng.
" " Math.
Per Centile Rank
Score
100 90 80 70 60 50 40 30 20 10 0
5 6 7 8 9 10 11 12 13 14 15 16

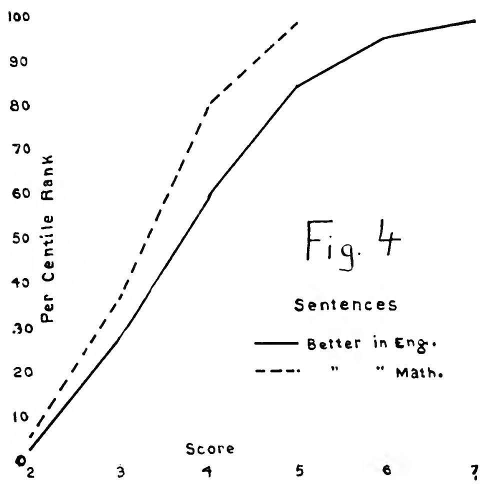


Fig. 4
Sentences
Better in Eng.
" " Math.
Per Centile Rank
Score
100 90 80 70 60 50 40 30 20 10 0
2 3 4 5 6 7

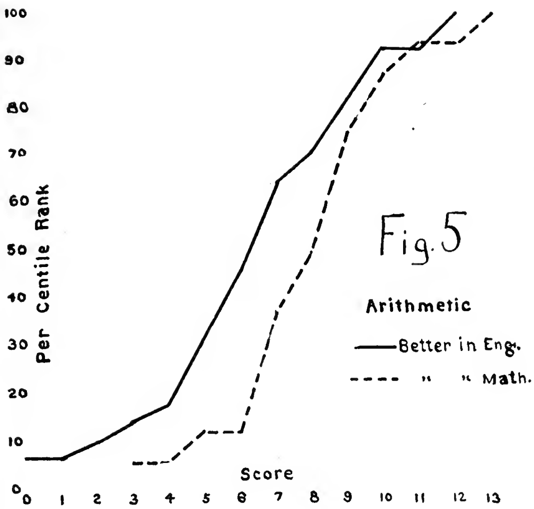


Fig. 5
Arithmetic
Better in Eng.
" " Math.
Per Centile Rank
Score
0 10 20 30 40 50 60 70 80 90 100
0 1 2 3 4 5 6 7 8 9 10 11 12 13

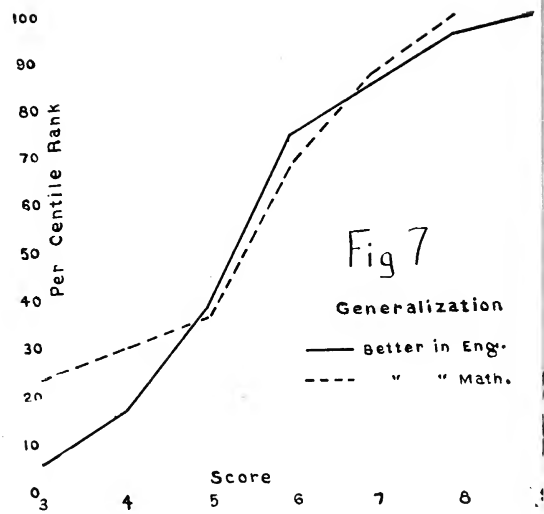


Fig 7
Generalization
Better in Eng.
" " Math.
Per Centile Rank
Score
0 10 20 30 40 50 60 70 80 90 100
3 4 5 6 7 8

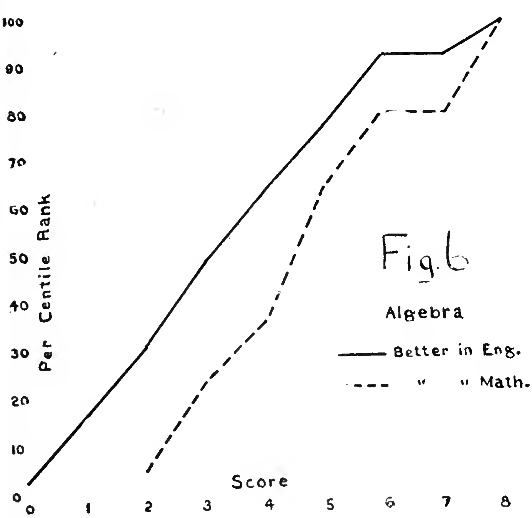


Fig. 6
Algebra
Better in Eng.
" " Math.
Per Centile Rank
Score
0 10 20 30 40 50 60 70 80 90 100
0 1 2 3 4 5 6 7 8

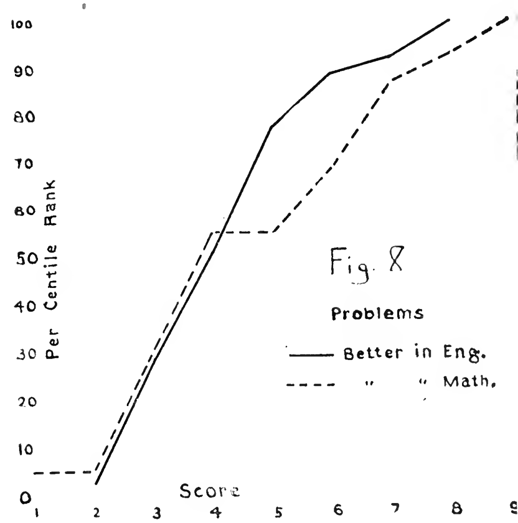


Fig. 8
Problems
Better in Eng.
" " Math.
Per Centile Rank
Score
0 10 20 30 40 50 60 70 80 90 100
1 2 3 4 5 6 7 8

coefficient of any size is that between the completion and the rhymes (.39). It will be observed, however, that the completion test correlates almost as highly as this with the three mathematics tests. It appears doubtful, therefore, if this correlation between it and rhymes has any significance as indicating a unique English ability common to it and rhymes alone. The correlation between arithmetic and algebra tests, on the contrary, is sufficiently higher than all the others to indicate very conclusively something common to these two tests not common to the English tests. One other correlation must be noted, viz., that of .42 between problems and sentences. It appears to be something of an anomaly. No explanation of it can here be suggested other than that perhaps both tests were of the nature of puzzles.

To sum up, the points of interest suggested by the table would be (1) The figures for the arithmetic and algebra tests bear out the evidence of the ogives that these are good mathematics tests. (2) The completion test would seem to be quite as much a general ability test as a specific English test. We may turn now to some other results we obtained which seem more interesting and suggestive.

It happened that 27 of the 43 subjects whose results are given above had taken both English and Mathematics as Freshmen at Northwestern, and that their grades were available. The grades were in the form of A, B, C, D, E, F. These were changed into numerical scores and averaged for the two semesters.[1] Table II presents the correlations between these two sets of grades and the 6 tests.

TABLE II

	English Grades	Math. Grades
Sentences	$.27 \pm .12$	$.10 \pm .13$
Rhymes	$.41 \pm .11$	$.14 \pm .13$
Completion	$.21 \pm .12$	$.36 \pm .11$
Arithmetic	$-.02 \pm .13$	$.57 \pm .09$
Algebra	$.15 \pm .13$	$.53 \pm .09$
Problems	$-.12 \pm .13$	$.14 \pm .13$

[1] In some cases the course had been dropped after a single semester, in which cases the grade for that semester alone had to be used.

Observing the three English tests, we again find that the completion test is of doubtful value for detecting specific English ability. For it here shows a higher correlation with mathematics grades than with English grades. The other two tests, however, show up better, especially the rhymes test. They correlate with English grades, and practically not at all with mathematics grades.

Turning to the three mathematics tests, the correlations in the case of the problems test are so small as probably to have little meaning. In the case of the arithmetic and algebra tests, on the other hand, the correlations with mathematics grades are decidedly high, while those with English grades are negligible.

We conclude, therefore, that the rhymes and sentences tests in so far as they correlated with the English grades, did so because of something in them which was not in English grades. These two pairs of tests seem, in short, to have brought to light specific and independent English and mathematics abilities.[1]

Two final questions arise. (1) Do the two tests of each pair measure the same thing, or has English ability two different phases, one of which is measured by the sentences test and the other by the rhymes test, and similarly has mathematics ability two different phases, one measured by the arithmetic test and the other by the algebra test? (2) Are the abilities which the tests measure merely a product of school training or are they something more fundamental? Neither of these questions can be answered with absolute certainty. Some considerations bearing on these points may be advanced, however. As regards the identity or non-identity of the factors measured, the evidence rests on the size of the intercorrelations between the two tests of each pair. In Table III we present these intercorrelations.

TABLE III

	English Grades		Math. Grades
Sentences..........	.27±.12	Arithmetic........	.57±.09
Rhymes...........	.41±.11	Algebra...........	.53±.09
Correlation of sentences with rhymes...........	.17±.11	Correlation of Arithmetic with Algebra.........	.50±.10

[1]The two sets of grades (English and mathematics) have an intercorrelation of .47±.10, indicating, therefore that these grades are an expression not only of those specific English and mathematics abilities, but also to a very large extent of a common something which we may dub "grade-getting ability."

The appearance of these results suggests that the sentences and rhymes tests measure different and relatively independent phases of English ability, whereas the arithmetic and algebra tests measure the same phase or nearly the same phase of mathematics ability. But to be sure of these conclusions we must resort to partial correlation.[1] See Table IV.

TABLE IV

Total Correlations		Partial Correlations	
Sentences with Eng. Grades	27±.12	Sentences with Eng. Grades (Rhymes constant)	.23±.12
Rhymes with Eng. Grades..	.41±.11	Rhymes with Eng. Grades.. (Sentences constant)	.39±.11
Arith. with Math. Grades...	.57±.09	Arith. with Math. Grades... (Algebra constant)	.42±.11
Algebra with Math. Grades	53±.09	Algebra with Math. Grades (Arith. constant)	.34± 12

It will be observed that eliminating rhymes has practically no effect upon the correlation between sentences and English grades, and likewise that eliminating sentences has practically no effect upon that between rhymes and English grades. These two tests are, therefore, correlated with English grades by virtue of measuring different things. It would seem, in other words, that English ability has at least two different phases, one of which is approximated by the sentences test,[2] and the other by the rhymes test.

Turning now to the mathematics tests, we find that eliminating algebra does appreciably lower the correlation of arithmetic with mathematic grades, and similarly, that eliminating arithmetic appreciably lowers the correlation of algebra with mathematics grades. We conclude, therefore, that for each of these mathematics tests part of the high correlation between them and mathematics grades is due to an element which is common to both tests. It is to be noted, however, that this common element does not explain the whole of the correlations. Even after it has been eliminated, the arithmetic test shows a correlation with mathematics grades of .42 and algebra test one of .34. To sum up: The sentences and rhymes tests seem to measure things mutually exclusive, both of which, however, correlate in part with English grades. The arithmetic and algebra tests, on the other hand, measure in addition to two such mutually exclusive things, correlating with mathematics grades, another common something which also correlates with mathematics grades.

We turn now to the last question. Are the things measured merely a product of school training or are they something more

[1] Yule, G. U. *An Introduction to the Theory of Statistics.* London, 1912.

[2] The size of the probable errors is such as to render the value for the sentences not absolutely conclusive. It is of course unfortunate that we did not have more subjects. We shall assume, however, that increasing the number of subjects would have reduced the probable errors for the sentences test without materially affecting the relative size of its coefficients with English grades and with mathematics grades.

fundamental? In other words, how far are the correlations of our tests with the grades to be explained simply and solely because the things measured in the tests were the things actually taught in the courses leading to the grades, or how far are they to be explained by assuming that the tests measured something more fundamental, *i. e.*, innate or partially innate abilities upon which the school successes were themselves dependent? Comparing our four tests from that point of view, the probability would seem to be that success in the two English tests depended more upon innate abilities in this sense, and less upon formal school training than was the case in the two mathematics tests. Speed in getting rhymes and in mentally sorting groups of words to make coherent sentences obviously are not directly taught in either school or college. Mental arithmetic and the handling of algebraic symbols, on the other hand, are. The rhymes and sentences tests seem, therefore, to get at something more fundamental than the two mathematics tests. In connection with the latter, however, a further suggestion presents itself. May it not have been that that which was seen to be common to both tests and to mathematics grades was the effect of schooltraining; whereas the factors which were in each test independent of the other and remained in the partial correlations were of the nature of innate and fundamental abilities? Such a possibility offers a suggestion for further study.

Finally, before concluding, the rhymes test and the algebra test may be mentioned as of special interest; the former because it is so very simple and yet gave a correlation of .41 with English grades; and the latter because it seems new and depends for success, we believe, rather upon the ability to do a particular kind of mathematical reasoning than upon any practise in the use of the algebraic symbols as such.

In conclusion, the following points seem worth summing up: All but a small percentage of an unselected group of college students were able on a subjective basis to class themselves as better in either English or mathematics. The ogives from a number of simple tests chosen arbitrarily as either English tests or mathematics tests agreed with this classification. Furthermore, a pair of these tests, rhymes and sentences, was found which correlated with English grades but not with mathematics grades, and another pair, arithmetic and algebra, which showed just the opposite correlation. The two phases of English ability measured by the sentences and rhymes tests seem to be independent of one another. The two phases of mathematics ability, on the other hand, measured by the arithmetic and algebra tests were only in part independent. The rhymes and sentences tests probably did not measure formal school training at all, and the arithmetic and algebra tests probably depended for success only in part upon such formal training. The rhymes and algebra tests are worthy of further study.

VOCABULARY OF GRAMMAR GRADE SCHOOL CHILDREN

Considerable interest has been manifested in recent years in the study of vocabularies. It has been fairly well established that the writing vocabulary of the average eighth grade pupil is about 2100 words. The writing vocabulary of the average adult is 3200 words. What is the extent of the reading vocabulary of the grammar grade pupil?

With a view toward determining the extent of the reading vocabulary of grammar grade children by using an easily operated plan for approximating the vocabulary, a study of 125 children in grades five, six, seven and eight of the public schools of Charles City, Iowa, was undertaken by the writer. The pupils were equally distributed throughout the four grades, and well represented a random selection from the four hundred fifty children in the four grades of the city.

From Webster's Elementary School Dictionary one hundred words were selected at random. These words were given orally to the one hundred twenty five pupils. Each pupil was given an opportunity privately to give orally at least one definition of as many of the hundred words as he was able. He was credited with knowing the meaning of the word if he could give one of its meanings, or use the word correctly in a sentence.

After an interval of four weeks these words were given again. They were arranged on typewritten sheets and each pupil was given an opportunity to mark each word as known, unknown, or uncertain. All words marked uncertain were used in a sentence and credit was given for all correctly used. This test is referred to as the Checked Word Test.

A second type of written measurement was taken by arranging four definitions for each word, three of which were incorrect and one was correct. The pupils were asked to check under each word the correct definition. Both of the written tests were given to the pupils in groups of thirty. Sufficient time elapsed between each test so that none of the words were remembered and there was no indication that any pupil had taken the time to look up the definitions of any whose meaning was not recognized.

Table 1.

Table 1 shows the distribution of median number of words known from the one hundred.

Grade	V	VI	VII	VIII
In the Oral Test	43	53	62	72
In the Written Test	50	51	65	74
In the False Definition Test	50	58	63	72
Average of the three	46	54	63	73

Table I is read as follows: in grade V in the oral test the middle or median pupil knew 43 of the one hundred words; in the eighth grade 72 words were known by the median pupil.

Basing the determination of the size of the vocabulary upon the oral test, Table II shows the number of words in the reading vocabulary of the median pupil in each grade.

Table II

Grade	V	VI	VII	VIII
Vocabulary Index	43	53	62	72
Median, Age-Yrs.-Mo.	10-11	12-1	13-0	14-5
Median Vocabulary	14,448	17,808	20,830	24,192
First Quartile	12,768	14,448	17,808	21,504
Third Quartile	17,808	20,024	21,840	25,200
Quartile Deviation	2,520	2,788	2,016	1,844
Range from lowest	6,720	11,088	11,424	15,458
to highest	21,504	24,528	26,208	27,216

Table II is read as follows: in the oral test the record of the middle pupil in the fifth grade, age 10 yrs and 11 months, is a reading vocabulary of 14,448 words. One fourth of the class has a reading vocabulary of 12,768 words or less, while one fourth of the class has a vocabulary of 17,808 words or more. One half the class ranges within 2,520 words of the median. The whole class ranges from 6,720 to 21,504 words. In a similar way the other columns are read.

It will be noticed from Table II that the progress from grade to grade is quite constant and averages about 3300 words.

In order to measure the range of vocabulary of any class, the task of giving an individual oral test is too great for practical use. The difference between the oral test and the written test in which a pupil is asked to mark words as unknown, known or uncertain is shown by reference to Table III.

Table III

Grade	V	VI	VII	VIII
Median Vocabulary Oral Test	14,448	17,106	20,830	24,192
Checked Word Test	16,870	17,136	21,840	24,864
False Definition Test	16,800	19,480	21,160	24,192

Table III shows rather striking similarity in results by the use of a marking method. In no case is the variation over 2400 words, or approximately 14%. In the Checked Word Test the average difference from the Oral Test is 1000 words or about five per cent. The greatest difference is in the fifth grade which shows a tendency to overestimate.

The above results as to the size of the vocabulary is based upon the calculated estimation that Webster's Elementary School Dictionary contains 33,600 commonly used words. From this list the one hundred words were selected at random.

The author claims that the dictionary contains 45,000 words. Using this number as the basis for calculation, the size of the vocabulary is increased 34 per cent, or approximately one third. The size of vocabulary on this basis is shown by Table IV.

Table IV

Grade	V	VI	VII	VIII
Oral Test	19,350	23,850	27,900	32,400
Checked Word Test	22,500	22,950	29,150	33,300
False Definition Test	22,500	26,100	28,350	32,400

Table IV is read as follows: the reading vocabulary of the median pupil in the eighth grade was 32,400 in the oral test. It was the same in the false definition test and 33,300 in the checked word test. Similarly the other grades are read.

Doctor Daniel Starch, University of Wisconsin, has determined the extent of vocabularies of children in grades five, six, seven and eight. Dr. Starch used the Webster's Unabridged Dictionary. A comparison between the results secured by the use of the Starch English Vocabulary Test and tests previously described in this article is shown in Table V.

Table V

Grade	V	VI	VII	VIII
Starch Test	21,840	27,040	31,200	41,600
Starch Standards	27,040	31,200	34,320	38,480
Elementary School Dict. Oral Test Vocabulary	19,350	23,850	27,900	32,400

Table V is read as follows: in grade V the size of the vocabulary of these pupils as determined by the Starch Test was 21,840, which is 5200 words below the standard determined by Dr. Starch. The size of the vocabulary determined by an oral test using Webster's Elementary Dictionary as previously explained is 19,350 words, a difference of only 2,490 words. In the same manner the figures for the other grades are read.

It is seen from the above discussion that the extent of a pupil's vocabulary will vary according to the dictionary used and whether the actual number of words claimed by the authors of the dictionaries or the number of words determined by counting are used. The difference in vocabulary size made by the use of a large or small dictionary is small. The average difference with these pupils is 4500 words. The actual difference in each grade is less than 3000 with the exception of the eighth grade which shows a difference of 9000 words.

The result of this study seems to indicate that an average pupil in fifth grade, age eleven years, has at least a reading vocabulary of about 15,000 words. It may vary as much as 6000 words according to these measurements or 12000 words according to the standards set by Dr. Starch, reaching 21000 or 27000 words. An average sixth grade pupil, age twelve years, knows at least 18000 words and ranges as high as 31000. An average seventh grade pupil, age thirteen, has a minimum vocabulary of 21,000 words and may range as high as 34,000 words. An eighth grade pupil will have the use of at least 24000 words and may know as high as 38000.

It is not claimed that these figures are final. It is believed, however, that they do show results sufficiently reliable to be indicative and significant.

After this manuscript was prepared, Professor Seashore reports having used the above as an information test in the selection of radio operators with the following two changes:

First: The examiner took each paper in the order in which the men were ready and marked the two most difficult words that had been checked in each of the four columns. He then required the person tested to use each of these eight words correctly in a sentence. For each failure, two points of merit were deducted. He thinks that this affords a satisfactory control of the checking particularly if internal evidences are taken into account. The test then becomes a test of "writing vocabulary."

The second change was to substitute the words, fa, friar and olivoid, for the three proper names in the original list, for evident reasons.

F. T. VASEY,

Superintendent of Schools,
Mason City, Iowa.

TRANSFER AND INTERFERENCE IN CARD-DISTRIBUTING

The object of this study was to discover the transfer and interference effects in card-distributing. The experiment consisted in distributing 150 cards into the 30 compartments of a box. The compartments were arranged in six rows of five each, and were numbered from 11 to 40. The cards were numbered correspondingly, five of each number. The compartments were not numbered consecutively, but quite without system. Before the beginning of each experiment the cards were always thoroughly shuffled.

The subjects were eight University women. They worked in two groups of four each. Practice was for approximately one hour daily. Part of the time was consumed in collecting and shuffling the cards.

The compartments were numbered in two ways, which we shall call scheme 1 and scheme 2. In scheme 2 the same numbers were used but *applied to different compartments*, and entirely without system as in scheme 1.

Group A alternated from scheme 1 to scheme 2 from day to day throughout the experiment, 30 days. Group B used scheme 1 only, for fifteen days, and then scheme 2 only.

We shall first notice the transfer effects. In figure I the results of the whole experiment are shown graphically. The vertical axis represents seconds; the horizontal, days. The practice curves are plotted from the daily averages of each of the two groups.

On the 16th day group B began to use scheme 2. Their average distribution time on this day was only 42.8% of the average time of distribution on the first day with the first scheme. These results show that in spite of the inhibitory effects of the habits established in the first fifteen days, efficiency in scheme 2 was much greater the first day than it had been the first day with scheme 1. On the fourth day with scheme 2 group B lacked only 3 seconds making the time attained on the fifteenth day in the first scheme.

In figure II are shown the results of the first five practices of group B with schemes 1 and 2. In scheme 2 the records are very much better than in scheme 1, showing not only rapid initial speed but also rapid improvement.

It is clear from this experiment that in distributing cards one gains efficiency in mastering such situations. After building up a system of habits for one scheme, a subject can more quickly build up a system of habits for another scheme. What is the explanation? There are probably several factors involved. In the first place, in our experiment the two systems of habits are not entirely different. The stimuli are the same, and the responses are the same, but the stimuli are coupled with different responses in the two cases. Whatever facilitation there is in the passage of nerve impulses through the neurones involved, arising from practice in the first experiment, is available and effective in the second experiment. Only *one* new nerve coupling has to be made for each card-number in the second experiment. This is more easily made than all the couplings originally involved.

In the first experiment the subject acquires skill in perceiving and recognizing the card-numbers; in getting the cards out of the pack; in getting them to the appropriate compartments. All these skills are effective in the second experiment. The transfer effect here is great, but entirely explainable from the standpoint of identical elements involved.

A comparison of A and B in figure 1 shows the inhibitory effects of one set of habits on the formation of the other. Group A alternated from one scheme to the other from day to day. If we may judge from the first day's work, they were faster learners than group B, but the efficiency which they acquired was much less.

Group A's final speed was 10% slower than group B's in scheme 1, and 22% slower in scheme 2.

The comparison can be made in another way: in the case of group B the final speed in the two schemes averaged 19% of first day speed, while in the case of group A the average final speed for the two schemes was 24% of their first day speed. Therefore, whether we compare absolute speed attained, or compare the absolute final speed with initial time, the result is the same. The group that alternated from one scheme to the other from day to day was at a great disadvantage in its method.

The inhibitory effects are further shown by reference to figure III. In the first fifteen days of the experiment, as shown in figure I, group B learns very much faster than group A, but at any time up to the fifteenth day, group B has had twice as much practice on scheme 1 as group A has had, for group A is also perfecting another set of habits; namely, those involved in scheme 2. In figure III are shown the records for group B for the first fifteen days, and all the records of group A for scheme 1, *i. e.*, the records on alternate days for thirty days. At any time after the first day's record, therefore, group A has had twice as much practice in distributing cards as has group B. In spite of this fact, however, owing to the inhibition of scheme 2 on scheme 1, group A does not attain the skill in scheme 1 attained by group B.

The inference from this experiment is that it is not economical to form at the same time two mutually inhibitory sets of habits. The better procedure is to form one, and then the other.

University of Missouri. W. H. PYLE.

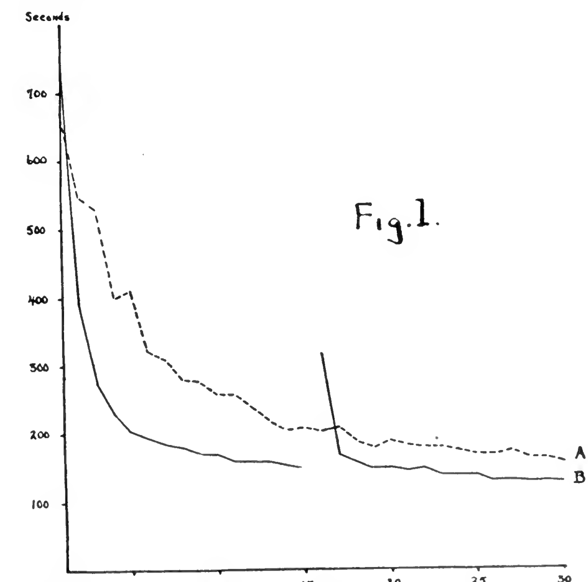


Figure 1. B shows the record for the group distributing 15 days with scheme 1, then 15 days with scheme 2. A shows records of group alternating between the two schemes from day to day for 30 days.

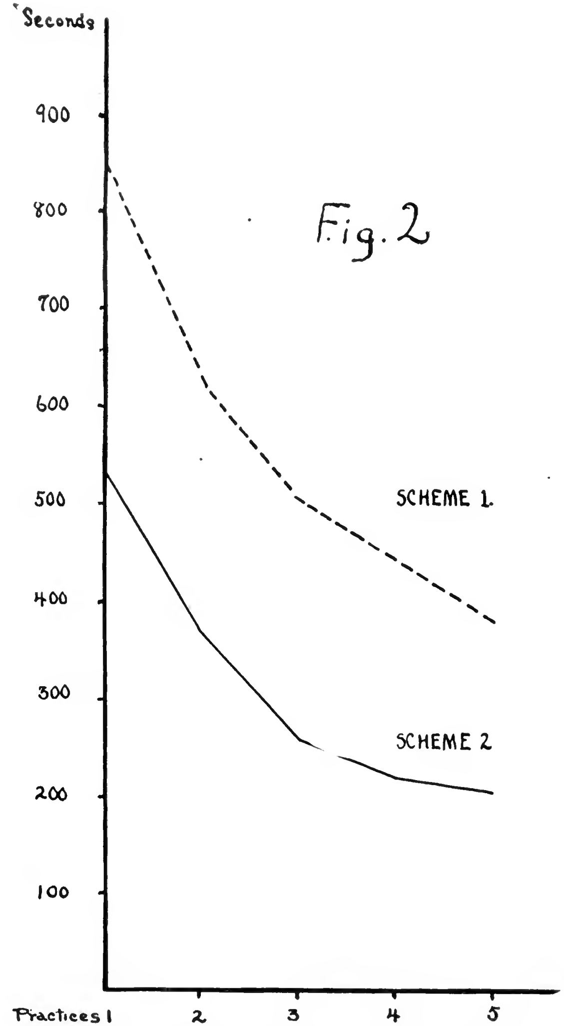


Figure 2. These graphs show the records for the first five distributions of group B with schemes 1 and 2. The effects from scheme 1 greatly reduces the time in scheme 2.

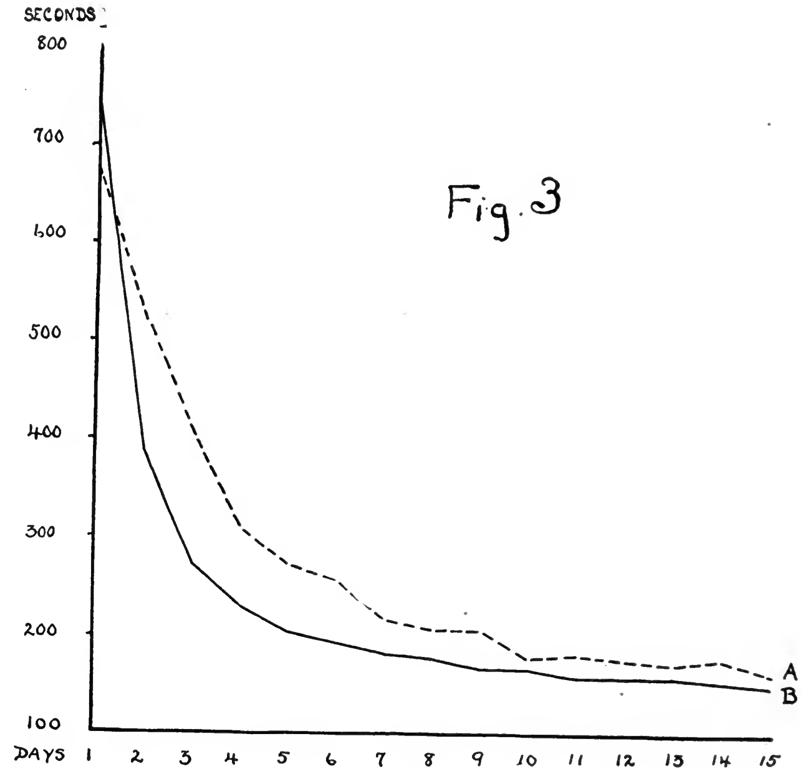


Figure 3. B shows the speed records for the group distributing daily according to scheme 1. A represents the records for scheme 1, of group A, distributing over a period of 30 days on alternate days. At any point after the first day, group A has had an equal experience with group B in scheme 1, and in addition, an equal amount of practice in scheme 2.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY
AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except July and August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated.

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of ten monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X FEBRUARY, 1919 NUMBER 2

EDITORIAL

EDUCATIONAL MEASUREMENTS AND THE CLASSROOM TEACHER

Educational measurements may be used in many ways for the evaluation of the work of the schools. Three of these uses may be singled out for special notice. First, they may be used by a commission or other outside body appointed to investigate and report on the efficiency of the school organization and instruction. This is illustrated in many of the school surveys that have been made in the last decade, and that have probably done more than anything else to spread abroad the conception of standard educational scales. A noteworthy early instance was the use of the Courtis Standard Tests in Arithmetic in the investigation of the New York School Inquiry Commission. The report on the application of these tests in New York City is a contribution to the scientific study of education which has never received the attention that it deserves. The use of educational measurements in this manner has unquestionably justified itself. It has enabled the survey commissions to get away from casual opinions and to base their recommendations on objec-

tively verified facts. It has given us results from widely separated communities which are gradually crystallizing into norms of performance. On the other hand these raw scores fail to give a faithful picture of the school situation, and need to be interpreted and explained in the light of other considerations which are frequently ignored.

A second use of educational measurements is seen in the activities of the special bureaus for educational research that are being established in an increasing number of school systems. It is the function of such a bureau to keep up a sort of continuous survey from within the system, and its influence is unquestionably more salutary and lasting than the transient survey from without. The directors of these bureaus are among our foremost students in educational research, and are gradually accumulating a mass of data which is already having its effect upon the organization and administration of schools, and which is bound to be of increasing significance in the future. Many practical school men take the view that educational measurements are so technical and require such a high degree of training that they ought to be left entirely to the expert specialists of such bureaus. Without ignoring the need for many more such specialists in the schools than we have at present, and without derogating from the splendid work that research bureaus are doing, it may be contended that this is not the chief service that educational measurements may render the schools. So long as measurements are regarded as the province of highly trained technicians the results will be kept at a respectful distance by principals and teachers.

The third and most beneficial use of educational measurements will be that by the principal and classroom teacher as a regular part of the school routine. We must disabuse our minds of the notion that measurements are something esoteric and abstruse. There is not a scale or a test now before the public that cannot be given, scored and tabulated by an intelligent teacher with a little careful practice. Measurements, instead of being looked upon by the teacher as the mysterious functions of a remote bureau, will come to be regarded as the teacher's useful tools, whereby she can determine quickly and accurately the needs of her pupils and the most effective adjustment of her energies. To this end many of our directors of research bureaus are making noteworthy contributions. Instead of more work for the teacher measurements mean in the long run less work, and much easier and more successful work. The recognition of the value of measurements for the teacher can only come about through the encouragement of the supervising staff. When superintendent, principal and teacher unite whole-heartedly to study the needs and capacities of individual children and to base the school work upon their findings, much of the drudgery and grind of teaching will disappear, and educational measurements will attain their greatest usefulness.

J. C. B.

NOTES AND NEWS

Dr. C. Truman Gray, of the University of Texas, has recently issued his Standard Score Card for Measuring Handwriting in four different forms. Form 1 is intended for training purposes, Form 2 has spaces for recording the score for each of the ten months of the school year, Form 3 is 17″ by 22″ in size and is intended to be hung on the wall of the school room for class use, and Form 4 is to be used as a report card to parents. These cards will be supplied to teachers at cost plus postage. At present Forms 1 and 2 cost fifty cents per hundred, Form 3 ten cents each, and Form 4 thirty cents per hundred.

The Frasier Writing Scale has recently been issued by the Department of Education of the State Normal School at Cheney, Washington. It is constructed on the basis of the Thorndike Scale from samples collected from the fifth, sixth, seventh and eighth grades of the Spokane Schools. The value of each step is given in terms of the Thorndike Scale, and norms for each grade are indicated, based on Freeman's standards. Advantages claimed for the scale are that the samples are all of slant writing, the same text is used in each sample, the samples are on lined paper like ordinary writing, and the scale is arranged with especial reference to its use by children.

Superintendent Ernest C. Witham, of Southington, Ct., has published a new series of Standard Geography Tests, The United States. These are issued in convenient form for class use, and contain nine tests, including an outline map of the United States, its bounding countries, native state and its capitol, ten rivers, five mountain ranges, five bodies of water, fifteen cities, five states, and ten industrial regions. The test is intended for the sixth grade or above, and the author supplies directions for giving and scoring the test, class record sheets, and blanks for graphs.

Dr. S. L. Pressey, of Indiana University, has devised two interesting series of "cross-out" tests for mental survey work. Each consists of four tests, as follows: Series 1, Primer Scale, I. Dot pattern test (in each group cross out the "extra" dot); II. Classification test (in each square cross out the thing that is different from the other two); III. Form board puzzle. (in each square cross out the piece you will have left over after you have fitted all the other pieces into the four forms at the top of the page); IV. Absurdities (in each square cross out the part of the picture that is wrong). Series 2, I. Verbal ingenuity (disjointed sentences containing one extra word to be crossed out); II. Logical judgment (in each list of five words cross out the thing that doesn't belong with the other four); III. Arithmetical ingenuity (in each set the

numbers are arranged according to a rule, but there is one number that breaks the rule. Cross it out); IV. Moral judgment (in each list of five acts cross out the one that is worst). The reactions of school children to these tests should furnish very interesting material. The tests are planned for the comparison of classes and schools as to the average of distribution of the abilities of the children, as to the classification of children according to ability, and as a check on other tests. Each scale takes twenty-five minutes to give, and can be scored in less than a minute.

Dr. L. W. Sackett, of the University of Texas, has issued his Scale in Ancient History in convenient booklet form, with blank spaces for the pupil's personal report, direction for giving the tests, the scale values opposite each question, directions for scoring, and June standards for each grade and sex in each test. Dr. Sackett is also working on a new form of United States History test which he expects to have ready by next fall.

Among the investigations in progress by the University of Ilinois Bureau of Educational Research, Dr. B. R. Buckingham, Director, are the derivation of the following scales: A scale for reasoning problems in arithmetic, a geography scale for the seventh and eighth grades, a history scale for the seventh and eighth grades, an English grammar scale for the same grades, standard requirements for memorizing literary material, and a group mental test for young children.

Mr. S. A. Courtis, director of educational research in the Detroit public schools, has issued a series of standard supervisory tests designed to further the use of tests as a routine school measure. The tests include spelling, arithmetic, geography and writing, and the descriptive folder of 52 pages contains in addition to much other valuable matter a percentage table and a table of rate-interval scores. With each test go individual and class record cards, and duplicate record sheets, half to be kept by the teacher and the other half to be sent to the principal or superintendent. On these record sheets the scores appear in five groups, children of standard ability who need no drill, those for whom the regular work will furnish sufficient drill, those in need of thorough drill, those needing special attention and extra drill, and those requiring special adjustment of work. While the bookkeeping looks somewhat formidable, there is little doubt that teachers who take and use such records will be able to do much more effective work than is usually attained.

The Adjutant General's Office of the War Department announces that a history of personnel work is now being prepared and will shortly be published in two volumes under the general title "The

Personnel System of the United States Army." Volume I, "The Evolution of the Personnel System," gives a history of the development of the work; Volume II, "The Personnel Manual," contains the operating instructions for the system as it existed on November 11, 1918. Further information regarding this publication may be obtained from Lt. Col. E. K. Strong, Jr., 528 State, War and Navy Building, Washington, D. C.

The general topic of the January meeting of the New York Society for the Experimental Study of Education was "Problems in Science Teaching." Professor Maurice A. Bigelow, of Teachers College, discussing "Problems Concerning the Content of General Science," raised questions in regard to the fundamental purposes of general science, to the content of the course, and to the best methods of approach. Dr. Otis W. Caldwell, of the Lincoln School, gave an account of the "Gary Science Survey," describing some very interesting tests used in it to determine whether "science trains in the observation and discrimination of things worth while." Principal John E. Wade described experiments in the teaching of science in elementary schools, and Mr. John McCarthy, Principal of the Singer Evening School, spoke of science in relation to shop needs.

Among the many interesting papers on the program of the meetings of the Department of Superintendence and affiliated societies at Chicago, February 24–28, we note the following: "The Relation between Study and Reading" by William S. Gray, University of Chicago, "Intelligence Testing as an Aid to Supervision", by Theodore Saam, Council Bluffs, Iowa; "Procedure with Standard Tests in the Supervision of English Teaching" by C. C. Certain, Detroit, Mich.; "Some Measurements of a Class of Children of Superior Intelligence" by Henrietta V. Race, Louisville, Ky.; "Results of Supervised Study in English as Measured by the Harvard-Newton Scale" by Charles Fordyce, University of Nebraska; "Measuring the Effects of Supervision in Geography" by S. A. Courtis, Detroit; "Variation in the Spelling of Root and Derived Forms of Words" by Ernest J. Ashbaugh, University of Iowa; "The Reading Problem in the Public Schools, as Affected by Actual Measurements" by Joseph P. O'Hern, Rochester, N. Y.; "Diagnosis as a Result of Measurement in Fractions" by Arthur W. Kallom, Boston; "Psychological Service in the Army and its Meaning for Educational Purposes" by Major M. E. Haggerty; "Educational Research from the Point of View of the Superintendent" by C. E. Chadsey, Detroit; "Organization for Effective Research" by Wilford E. Talbert, Carnegie Institute of Technology; "Citizen Co-operation in Educational Research" by F. E. Shapleigh, Buffalo, N. Y.; "What Educational Results of the Kindergarten may and may not be Measured?" by Alice Temple, University of Chicago, Julia Wade Abbott, Washington, D. C., W. C. Bagley, Teachers College, and Lotus D. Coffman, University of Missouri.

The National Society for the Study of Education has a Committee on the Materials of Instruction, Professor C. H. Judd, University of Chicago, chairman, which is endeavoring to collect materials and devices now being successfully used by individual teachers, and to make these more widely known and generally accessible. The committee is concentrating its attention on grades seven to nine, and is seeking information on books and materials used, on original work by individual teachers, and on plans for organizing correlated courses in these grades.

Dating from the present semester every course in the freshman and sophomore years of Yale University will be required. The required courses include economics, philosophy and American and European history.

Professor H. Austin Aikins recently gave two lectures at the Carnegie Institute of Technology, Pittsburgh, before the occupational therapy class on problems of psychotherapy and the work in Canadian hospitals of importance for occupational aids. At the same institution Lt. Col. E. K. Strong, Jr., spoke on job analysis before the Research Bureau for Retail Training.

At the Baltimore Meeting of Section L (Education) of the Association for the Advancement of Science Professor V. A. C. Henmon, of the University of Wisconsin, was elected vice-president.

Miss Lida Lee Tall, supervisor of grammar grades of Baltimore, Md., has been appointed supervising principal of the first six grades at the Lincoln School, New York City.

PUBLICATIONS RECEIVED

WILLIAM H. ALLEN. *War Fact Tests for Graduation and Promotion.* Yonkers-on-Hudson: World Book Company, 1918. Pp. 80. Twenty-four cents.

A pithy statement of the reasons why we are at war, our peace aims, home town war facts, home state war facts, home country war facts, world war facts and after-the-war needs. The home town and state war facts are left somewhat schematic, and directions are given for securing and inserting the appropriate data. The book has been approved for use in New York City schools.

F. MATTHIAS ALEXANDER. *Man's Supreme Inheritance. Conscious Guidance and Control in Relation to Human Evolution in Civilization.* New York: E. P. Dutton and Company, 1918. Pp. xvii, 354. $2.00.

This is a thesis on the importance of correct breathing. The author has much to say comparing modern conditions of life with the habits of primitive man, and both on general grounds and on the basis of specific cases urges the importance of "respiratory re-education." Incidentally he voices many good ideas about "the open mind," and the value of a scientific point of view. The book has a commendatory introduction by John Dewey.

FRANCIS G. ALLINSON, Editor. *The Aeneid of Virgil.* In English translation by John Connington. New York: Scott, Foresman and Company, 1916. Pp. 452. Forty cents.

One of the strongest arguments in favor of the Greek and Latin languages has been the educative influence of contact with the masterpieces of their literatures. But modern educators are coming to believe that this result can be better attained, with much saving of time and energy, through translations. It is therefore, with satisfaction that we note the issue of this school edition of the best known prose translation of Virgil's Aeneid. We believe that young people can be gotten to read this classic with real interest and enthusiasm, and that in three months' time a vastly better idea can be gained of the great Latin epic than from a three years' study of the Latin language.

FRANK P. BACHMAN. *Great Inventors and Their Inventions.* New York: American Book Company, 1918. Pp. 272. $.80.

Twelve stories of great inventions, with a concluding chapter on the famous inventors of to-day. There are accounts of the development of steam and electric power; of such manufactures as the cotton gin, spinning machines, sewing machines, reaping machines and making of steel; means of communication, as printing, telegraph, and telephone; and recent inventions, as the electric light, flying machines, wireless telegraphy, and submarine boats. This is really educative reading material.

LYDIA RAY BALDERSTON. *Housewifery: A Manual and Text Book of Practical Housekeeping.* Philadelphia: J. B. Lippincott Company, 1918. Pp. 353. $2.00.

This is another of the excellent Lippincott's Home Manuals. In it the previous high standard of usefulness, scientific accuracy, and mechanical perfection has been worthily maintained.

MAURICE BARRES. *Les Traits Eternels de la France.* New Haven: Yale University Press, 1918. Pp. 90. $1.00.

An address delivered before the British Academy in the hall of the Royal Society, July 12, 1916, in which the noted French Academician pays tribute to the character of the French soldier and to his loyalty and devotion to France.

ARTHUR L. BEELEY. *An Euperimental Study in Left-Handedness, with Practical Suggestions for Schoolroom Tests.* Chicago: University of Chicago Department of Education, 1918. Pp. viii, 69. $0.50.

The uniformity of action demanded in ordinary school procedure makes the position of the left-handed child awkward and embarrassing. Shall the child be taught to prefer the use of the right hand in acts of skill, as do normally right-handed children? At present we have only masses of conflicting opinion. At the very outset we are confronted with the difficulty of diagnosing left-handedness, and of determining whether it is inherited or acquired. Extended experiments with the Jones brachiometer convinced the author that this instrument is absolutely unreliable as a test of left-handedness, and the most effective means of diagnosing the condition is offered by the tapping test and a tracing test devised by the author. The susceptibility of left-handed children to "mirror-writing" is due to the dominance of motor over visual controls of movement. The monograph closes with a concise summary of the literature on left-handedness, and a bibliography of 36 titles.

MILTON BENNION. *Citizenship: An Introduction to Social Ethics.* Yonkers-on-the-Hudson: World Book Company, 1917. Pp. xviii, 181.

The times are ripe for a new statement of the fundamental principles of right living, and the author of this book has rendered a service in proposing for the benefit of high school pupils a series of inquiries as to what we should do about many of the civic and social problems that confront us daily. There is a list of questions on each chapter.

J. D. BERESFORD AND KENNETH RICHMOND. *W. E. Ford: A Biography.* New York: George H. Doran Company, 1917. Pp. 318. $1.35.

A sympathetic and appreciative account of an English teacher and philosophical thinker who tried to apply the ideals of Comenius and Pestalozzi to the conditions of modern life. The book is denominated "the story of a man who lived before his time," but to the reviewer it would seem rather the tragedy of a sensitive spirit who neglected to make use of the methods of advertising which the age requires. It is to be hoped that the book will help to awaken the public in both England and America to demand greater sincerity and more vital earnestness in education.

LEE BYRNE, EDITOR. *The Syntax of High School Latin.* Chicago: The University of Chicago Press, 1918. Pp. xi, 60. $0.75.

The volume contains an interesting analysis of the mental processes involved in reading simple passages of Latin prose, and presents detailed statistics of the frequency of different types of constructions in the selections usually read in high school Latin. There follow pertinent suggestions on the relative time and emphasis to put on various phases of syntax in the material of the course. The book is invaluable to the teacher of Latin.

WILL DURANT. *Philosophy and the Social Problem.* New York: The Macmillan Company, 1917. Pp. x, 272. $1.50.

By the social problem the author means the problem of reducing human misery by modifying social institutions. Reflecting on the upheaval in Russia, the increase of interest in socialism all over the world, and the probable readjustments that we shall be called upon to face after the war, we readily grant the tremendous significance of this problem. The author believes that the solution is to be found only through philosophy, and devotes two-thirds of the book to an interpretation of the contributions of Socrates, Plato, Bacon, Spinoza and Nietzsche to the betterment of society. The remainder of the book is a plea for the dignity and importance of philosophical thinking at the present time. His plan is to organize the intelligence of the world to improve life, just as both the Germans and the Allies have organized the material and spiritual resources of the world to destroy life. It is an admirable scheme, and worthy of wide-spread consideration.

GEORGE ELIOT. *The Mill on the Floss.* Edited for school use by C. H. Ward. New York: Scott, Foresman and Company, 1916. Pp. 501. Forty-five cents.

This, the most recent volume of the Lake English Classics, is not overburdened with notes and allows the reader to maintain uninterrupted contact with the story.

N. L. ENGELHARDT. *A School Building Program for Cities.* New York: Teachers College Contributions to Education, 1918. Pp. ix, 130.

Most of our American cities have changed so rapidly in population in the past fifty years that a school building program has been made very difficult. Most cities have been satisfied to lag a goodly distance behind their building needs, and could scarcely be said to have any building program whatever. This volume presents systematic comparative studies of the populations of a group of medium-sized cities, considers the trend of practice regarding the details of the school plant, and gives valuable data on financing a building program. The investigation will be a boon to the superintendents of our city schools.

THOMAS E. FINNEGAN. *Teacher Training Agencies. A Historical Review of the Various Agencies of the State of New York Employed in Training and Preparing Teachers for the Public Schools of the State.* Albany: State Department of Education, 1917. Pp. 439.

An account of the Lancasterian schools, the training classes in the earlier academies, a history of each of the state normal schools, and city training schools, and a description of the facilities for training teachers in the other institutions of the state. The volume is sumptuously and profusely illustrated.

SIDNEY G. FISHER. *American Education.* Boston: Richard G. Badger, 1917. Pp. 171. $1.25.

This is a powerful and well-supported arraignment of present educational practices in America, with especial reference to secondary and higher education. The old cry that the American youth is two years behind his European fellow is revived with copious citations. Our schools pander to information and are too weak to furnish the rigorous disciplinary training that European schools provide. Teaching has become debased by catering to the interests and whims of the pupils. Our college faculties have abdicated in favor of the athletic coach, and the fraternity. Crowning crime of all our courses of study have been disorganized by the pernicious

elective system, and their entire effect is a mass of weakness, shilly-shallying, and inefficiency. It is a veritable Jeremiad, and will meet with much approval from those who lament the decay of the good, old times.

FLORENCE C. FOX. *The Fox Series of Readers. The Fox Primer from Mother Goose.* Pp. 170. *The Fox First Reader.* Pp. vii, 156. *The Fox Second Reader.* Pp. v, 187. *The Fox Manual for Teaching Reading.* Pp. viii, 62. New York: G. P. Putnam's Sons, 1918.

This attractive series of readers is constructed on the principle of socializing the recitation. In so far as possible direct discourse is used, and the narrative parts are so divided that they can be assigned to various pupils. In the primer much attention is given to phonic exercises and to oral story-telling.

FRANK N. FREEMAN. *The Handwriting Movement. A Study of the Motor Factors of Excellence in Penmanship.* Chicago: University of Chicago, Department of Education, 1918. Pp. xvi, 169. $1.25.

This monograph presents the results of an extended experimental investigation of writing movements carried on in the University of Chicago laboratory under a grant from the General Education Board. It is unquestionably the most thorough and elaborate study of writing movements yet made. There is an analysis of individual records from motion-picture studies, a statistical comparison of the position and movement of good and poor writers, and an account of a field study and a pedagogical experiment in writing. The author urges that the pupil be taught to take a critical attitude toward his writing, and that he be afforded definite knowledge of his standing and progress. Letters, he says, may be formed by such a combination of finger and arm movement as is most natural to the pupils.

GEORGE W. GERWIG. *Schools with a Perfect Score.* New York: The Macmillan Company, 1918. Pp. xiii, 194. $1.10.

This is not a statistical comparison of schools rated on the basis of standard tests, as one might surmise from the title. The author's aim is rather to picture in vivid fashion certain ideals of school grounds, school buildings, the activities of teachers, intellectual education, vocational education, and moral training. The book closes with a scathing arraignment of the defects of present school practices, and a summary of specific improvements to correct these defects.

WILLIAM ERNEST HOCKING. *Morale and its Enemies.* New Haven: Yale University Press, 1918. Pp. xvi, 200. $1.50.

"Morale, the invisible force behind war-making, is also the invisible force behind peace-making. It is the temper of a people expressing itself in action,—in war and in reconstruction. The conscious need of understanding and calling upon the high possibilities of that temper has never before been so recognized by a nation." This is one of the best philosophical analyses of the factors involved in the war. The author was director of the courses on Morale in the Northwestern Division of Army Camps during the summer of 1918, and has brought to bear on the problems of maintaining the war spirit the keen philosophical insight which has characterized his thinking in other fields.

VOLUME X MARCH, 1919 NUMBER 3

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

IS INDIVIDUAL LEARNING CAPACITY CONSTANT FOR DIFFERENT TYPES OF MATERIAL?

W. H. PYLE
University of Missouri

Learning is now defined as bond-building. In habit-formation, for example, nervous bonds or connections are established between stimulus and response. A nervous bond, of course, is merely a path of least resistance. In ideational learning bonds are established—paths of small resistance—between the cortical processes underlying the ideas involved. The question which we raise in this paper is this: are the characteristics of the nervous system such that a person may be spoken of as a "good learner" or a "poor learner;" is it easy for one person to establish new neural bonds, involved in learning, and relatively difficult for others? Must we say that a person can easily learn mathematics, but learns some other branch, let us say history, with difficulty? Or can we say that a person is a good learner or a poor learner of whatever he undertakes to learn? In the schools, for example, is a person known as a *good* student or a *poor* student in general, or is he known as a good student in history and language, for example, and a poor student in science and mathematics? Generally, the former is the case. Students are spoken of as good or poor without reference to any particular subject of study. There are, however, numerous exceptions. Sometimes a student is known as an excellent student in science and as a poor student in language. In such cases the question arises whether the student's being good in one subject and poor in another may be due to such factors as difference in the preliminary training for the subjects, or difference in interest in the two subjects. Several investigators have computed the correlation between the standing of University students in one subject with standing in another. These correlations have in all cases been found to be high, though of course not unity.

The question which we have raised is of the highest importance to education, for if learning, whatever its nature, depends upon fundamental characteristics of the brain and nervous system, then

it will be found a comparatively simple matter to determine by experiment the learning capacity of the children in the public schools and the students in the colleges. On the other hand, if learning is a specific matter, we cannot speak of learning capacity in a general sense, but must speak of ability to learn this or that.

The first experiments of the writer in this field were performed in 1911–1912. On the basis of these experiments the statement was made in *The Outlines of Educational Psychology* (p. 202) that "it seems probable that every individual has a fairly definite coefficient of learning capacity that is fairly constant." The few experiments conducted at that time consisted in having subjects commit to memory nonsense syllables. The lists contained twenty-six syllables each. The number of repetitions necessary to learn a list at one sitting was determined. The number of repetitions for re-learning a series on succeeding days was also determined. It was found that re-learning time had a very definite ratio to learning time. The experiments showed that quick learners retained well what they had learned, as determined by their later re-learning time. While the experiment threw no light on the matter of learning different types of material, it showed great definiteness and constancy in individuals in learning the same kind of material.

For several years I have been trying to discover how the matter stands with reference to different types of material. Ability in learning four different types of material has been studied; namely, substitution, nonsense syllables, card distribution, and marble distribution. These types of learning are quite different, one from the other. Substitution tests and nonsense-syllable learning are types of what may be called intellectual learning. The associations are in large measure between ideas. The motor element is unimportant in them; at any rate, is much less important than it is in the card and marble-distributing tests.

Three types of substitution tests were used; writing digits for symbols, symbols for digits, and symbols for the letters of the alphabet. In all of the substitution tests, except alphabet-symbol substitution, an experiment consisted of five minutes work. In some cases the experiment was continued for several five-minute periods. Four different keys were used.

In the learning of nonsense syllables the method was that of complete learning of a series of ten syllables each. The syllables

were exposed by means of a Jastrow tachistoscope, each syllable being exposed for a period of one second. After the exposure of the series the subjects wrote in their proper sequence all the syllables they could recall. The exposures were continued until all the syllables could be repeated in their proper order. The syllables were in all cases written down by the subjects immediately after the exposure. In computing the correlations we have in some cases considered the total number of repetitions required for the subject to get all the syllables; in some cases we used the efficiency after a certain exposure, say the second.

The card-distributing experiment consisted in distributing 150 cards into the 30 compartments of a box. The compartments were numbered from 11 to 40. The cards were numbered in the same manner, five of each number. The numbering of the boxes was without any system, *i. e.*, the number did not run consecutively on adjacent boxes. The cards were in all cases carefully shuffled before each distribution. Efficiency was of course determined by the time required for distribution.

The marble experiment consisted in distributing by means of a complicated machine, marbles of four different colors and two different sizes into their corresponding compartments. Efficiency was determined by the number of marbles correctly placed per minute.

The card and marble experiments are alike in both being motor learning, but they are quite different in detail. In the case of the card distribution, simple associations had to be built up between a number and a place. In the case of the marbles, the matter was much more complicated. Not only was it necessary to build up several types of associations, as between the color of a marble and the place where it was to be found, the color of a marble and the place where it was to be put, but also with the movements of the feet in operating pedals for different kinds of marbles. In the operation of the machine distinctions must constantly be made among the six different kinds of marbles, and four holes on the top of the table, and the movements of both feet and both hands. The experiment, therefore, demanded a high degree of co-ordination of muscular movements, and concentrated attention.

TABLE OF CORRELATIONS

The correlations which follow were all computed by the Pearson formula. The experiments on which they were based were performed at different times and upon different groups of students

Substitution (digit-symbol and symbol-digit) with:

		r.	P. E.
Cards		.64	.074
"		.66	.076
"		.559	.066
"		.59–	.09
"		.56	.06
"	Av.	.602	
Marbles		.469	.103
"		.54	.10
"		.62–	.058
"		.68	.036
"		.82	.072
"	Av.	.626	
Nonsense syl.		.432	.077
Alphabet-symbol		.57	.091
" "		.59	.08
" "	Av.	.58	
Average of substitution correlations			.594

Alphabet-symbol substitution test with:

		r.	P. E.
Cards		.54	.095
"		.49	.11
"	Av.	.515	
Substitution (digit symbol)		.57	.091
"		.59	.08
"	Av.	.58	
Average of alphabet-symbol correlations			.547

Nonsense syllables, with:

		r.	P. E.
Cards		.399	.08
Substitution		.432	.077
Mirror-writing		.505	.11
Marbles		.342	.084
"		.53	.196
	Av.	.441	
Average all nonsense syllable correlations			.441

Card-distributing with:

		r.	P. E.
Marbles		.31	.13
"		.218	.091
"	Av......	.264	
Substitution		.559	.066
"		.56	.06
"		.64	.074
"		.66	.076
"		.59	.09
"	Av......	.602	
Alphabet-symbol		.54	.095
" "		.49	.11
" "	Av....	.515	
Nonsense syllables		.399	.080
Average all card correlations			.496

Marble-distributing with:

		r.	P. E.
Substitution		.620	.058
"		.469	.103
"		.54	.10
"		.82	.072
"		.68	.036
"	Av......	.626	
Nonsense syllables		.342	.084
" "		.53	.196
" "	Av......	.436	
Cards		.31	.13
"		.218	.091
"	Av......	.264	
Average all marble correlations			.503

Average inter-correlation of all the learning tests. . .516

The ranking of the tests, based on their degree of correlation with the other tests, is as follows: digit substitution test, .594; alphabet substitution test, .547; marble distributing, .503; card distributing, .496; nonsense syllable learning, .441.

In the accompanying table the correlations are shown between each type of learning and the other types. These correlations are based upon experiments with several hundred University students, and were performed in the years 1914–1918. The groups studied vary all the way from nine to 120. The results from all of them are shown. It will be seen that the average of the "raw" correlations between one type of learning and the other types is approximately .50. The true correlation would be considerably higher. This study therefore shows a high, but not a perfect, correlation between ability for one type of learning and other types.

Two questions now arise. First, if all extraneous factors were eliminated, would the correlation be unity, showing that we can speak of learning capacity as a definite characteristic of the nervous system? Second, what is the cause of the positive correlation which we have actually found? These questions are really one, however, for if the true correlation between these types of learning is unity, there could be no question that the explanation would be a definite and constant learning capacity in the subjects.

There is considerable evidence that if all the disturbing factors could be removed, the true correlation would be much nearer unity than the raw correlation which we have actually found. The subjects were all mature men and women whose ages would average over twenty-one. Their experiences in life had been widely different; different prejudices and attitudes had been acquired. They therefore entered these experiments not only with different learning capacities, but with all sorts of differences due, not to learning capacity, but to past experience. In the marble-distributing experiment, for example, it was found that students who had had considerable experience in running typewriters, or in playing pianos, made better scores. In the card-distributing experiment not only was ability to learn the location of the boxes a factor, but facility in handling the cards was also a factor. In playing cards, some of the subjects had acquired great dexterity in handling them; others who had not played cards made poorer records because of lack of dexterity. And with each type of learning the subjects came to the experiment with differences due, not to native learning capacity, but to difference in past experience. Now, if taking the subjects just as we found them, without eliminating any of the disturbing factors, we found an inter-correlation of approximately .50, it is evident that

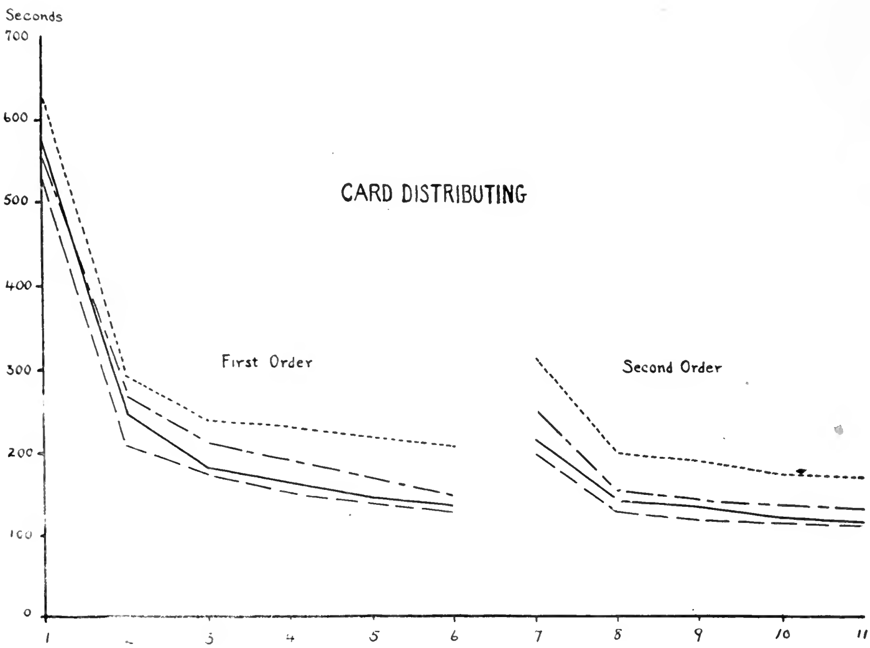
if these extraneous factors could be eliminated, the true correlation would be much higher. In the writer's opinion it would be unity. But of course this has not been demonstrated. It remains to eliminate all the other factors and find the relation of learning capacity itself in one kind of work with learning capacity in another.

However, there is another possibility. It may be that our assumption of a definite learning capacity in an individual for all types of material is erroneous, and that the positive correlation which we have found is due to the supposed "central" or "general" factor of the English school. The central factor may be *one* element in producing the positive correlation, but it can scarcely be the *sole* element. The inter-correlation in the following tests,—logical memory, rote memory, substitution, opposites, free association, word-building, and completion,—averages .231 as found by the writer in a careful study of sixty cases. If there is such a thing as central factor it certainly operates in all these seven tests. Still the inter-correlation is less than half that which we have found in the case of the four different types of learning.

How would this experiment have turned out if some of the learning had been in music, or in art, or in dancing? We cannot, of course, answer this question, because studies involving such widely different types of performances have not been made. It is probable, however, that skill in music, art, etc., involves not only general, but highly specific abilities as well. If this is true, and we should test out a number of people, finding their efficiency as learners in these different fields, any positive correlation found would be due to two factors; first, the degree in which the subjects possess the specific capacities involved, and second their degree of general learning capacity. The writer's view is that one has a brain in which it is easy or difficult to establish new bonds. Now in such a thing as music such other factors are involved as rhythm, distinguishing pitch, appreciation of harmony; and in painting, such specific abilities may be involved as type of imagery, space perception, etc. In a word, in most learning there are probably both general and specific factors, the former involving the bond-forming characteristic of the central nervous system, the latter involving other characteristics of the brain, sense organs, and muscles.

In the experiments reported here, especially in the card-distributing experiment which was, in several cases, continued over a period of 30 days, nothing was more noticeable than the definiteness

and regularity of the learning of the different subjects. They retained their relative positions or ranks from day to day with great constancy as, for example, in the card-distributing experiment as shown in the accompanying graph. At the beginning of an experiment the experimenter could predict with great accuracy the relative ranks of the subjects on the basis of their work on the previous day. The learning efficiency of the various subjects became definitely known to the experimenter. The characteristics underlying learning efficiency seem to persist and to be as definite as anything in nature. Such variations as occur in a given subject's learning from day to day have definite causes, and can usually be predicted when the causes are known to exist. Of course, constancy in learning always depends upon constancy of the conditions and factors involved. These factors are numerous. The condition of nearly every organ in the body can have its effect upon the temporary efficiency of the learner. But there is probably a constant factor which may be called general learning capacity, dependent upon the characteristics of the central nervous system.



MENTAL TESTS FOR COLLEGE ENTRANCE*

L. L. THURSTONE

Carnegie Institute of Technology, Pittsburgh

Purpose of the Mental Tests

One of the main duties of the division of applied psychology at Carnegie Institute of Technology is to advise the deans and the faculty concerning the admission of students and their subsequent promotion and transfer from one course of study to another. Our main concern so far has been to establish the diagnostic value of mental tests as a criterion for admission. In order to demonstrate the practical value of mental tests we have been giving a series of tests to the freshmen early in their first year of residence. The test papers were scored and filed until scholarship records and faculty estimates were available. Then we asked ourselves this question: If we had been given complete control over admissions could we have predicted by means of our tests which students were to fail and which were to succeed in their college work? This is a practical problem. Our results may be ever so interesting from some theoretical standpoint but even if we compile tons of norms of performance, the work is a failure if we are unable to establish a relation between the test scores and the criterion which is in this case the student's performance in college work. The present report is an answer to this question with reference to the entering students at the women's college at Carnegie Institute of Technology.

Description of Group

The students to whom we gave these tests were all freshmen at the Margaret Morrison Carnegie School of Carnegie Institute of Technology in Pittsburgh. All of the 114 freshmen were high school graduates. The four year course in the women's college is devoted mainly to applied science, including household economics, secretarial studies, costume economics, home arts and crafts, social work and general science. The freshmen look forward to specializing in one of the above mentioned departments. For this reason it is apparent that the school does not attract students of academic preferences, but the grade of work in the special subjects mentioned above is comparable with that of other well known women's colleges. The first year course, on which the instructors based their estimates of the students, contains the following subjects: physics, sewing, history, English, drawing and color, hygiene, chemistry, foods, accounting, and social ethics.

*Since this article was written our mental test program has been arranged in the cycle-omnibus form, using most of the test material here described. The results of the cycle-omnibus form will be published later.

Description of the Tests

I shall limit myself in this report to six of the mental tests which constitute our present standard series for the women's college at the Institute. These tests are as follows:

1. Completion Test

This is a selection from the Trabue Language Test material used by Professor Walter Dill Scott in the Bureau of Salesmanship Research at the Carnegie Institute of Technology.

2. Robinson Range of Interest Test

The instructions for the test are as follows:

Below is a list of statements. Some of these statements are true: some are false. In the ( ) to the left of each statement make a + sign if the statement is true, but a − sign if the statement is false. Even when in doubt make the mark which you think is most probably correct. Be sure to mark each statement. The first two statements are already marked correctly."

(+) 1. George Washington was the first president of the United States.
(−) 2. Christmas comes on July 4.

The following ten sample sentences illustrate the nature of the test.

( ) 3. A satisfactory rating in "Bradstreet's" is an important business asset.
( ) 8. A broken crank shaft generally means a ruined carburetor.
( ) 19. Arson is sometimes used as a medicine.
( ) 23. Sir Roger Casement was accused of treason and put to death.
( ) 49. Forbes Robertson's performance of "Hamlet" has thrilled many audiences.
( ) 51. The "squeeze play" calls for a long hit by the batter.
( ) 52. Stocks and bonds are sometimes offered as collateral.
( ) 69. The Iroquois fire would have been less serious if there had been an adequate supply of lifeboats.
( ) 101. John Marshall was a famous American jurist.

3. Opposites Test

This test has the following instructions:

"On the following page are forty pairs of words. If the two words of a pair mean the same, or nearly the same, draw a line under the word SAME. If they mean the opposite, or nearly the opposite, draw a line under the word OPPOSITE. If you cannot be sure, guess. The two samples are already marked as they should be."

fall–rise	same	*opposite*
class-group	*same*	opposite

"When the signal is given (wait for the signal) turn over this page and work as rapidly as you can without making mistakes."

The test proper has forty reactions like the samples in the instructions.

4. Gordon Directions Test

This test has four parts printed on the four pages of a pamphlet. Each part of the test has forty stimulus words. The instructions are as follows:

"Part 1. Work as quickly as you can without making mistakes. At the right of each of the following words make a figure. If a word contains the letter **e** mark it 1. If it does not contain **e** mark it 2. Look at the samples. Samples: Strike 1, train 2."

"Part 2: Work as quickly as you can without making mistakes. At the right of each of the following words make a figure. If a word contains the letter e and a mark it 1. If it contains the letter **e** but not **a** mark it 2. If it contains the letter a but not e mark it 3. Samples: change 1, every 2, calamity 3."

"Part 3: Work as quickly as you can without making mistakes. At the right of each of the following words make a figure. If a word contains the letters **e**, **a**, and **r**, mark it 1. If it contains the letter **e** but not **a** and **r**, mark it 2. If it contains the letter **a** but not **e** and **r** mark it 3. If it contains the letter **r** but not **e** mark it 4. Samples: dream 1, feeling 2, stand 3, stirrup 4."

"Part 4: Work as quickly as you can without making mistakes. At the right of each of the following words make a figure. If a word contains the letters **i**, **n** and **g**, mark it 1. If it contains **i** and **n** but not **g** mark it 2. If it contains **i** and **g** but not **n** mark it 3. If it contains **n** and **g** but not **i** mark it 4. If it contains i but not n and g mark it 5. Samples: ignite 1, thirteen 2, ligature 3, gnarled 4, ritual 5."

5. Analogies Test

The instructions for this test are as follows:

"On the following pages are forty lines of words. In each line the first two words are related in a certain way. Draw a line under the word in heavy type which has the same relation to the third that the second has to the first.

Notice the following samples:"

1. Foot-shoe; hand (head, *glove*, finger, clasp)
2. Mayor-city: captain (*ship*, private, general, town)
3. Rudder-ship: tail (*dog*, cat, sail, *bird*)
4. Penny-copper: dime (quarter, *silver*, nickel, spend)

The test proper contains forty problems.

6. Marble Statue Test

This memory test was given and scored according to instructions in Whipple's *Manual of Mental Tests*. Vol. II., p. 209.

Procedure in Giving the Tests

All the tests are printed in uniform size, 8½"x11." They are bound into a pamphlet, stitched and cut. One pamphlet is given to each subject. The tests are given to groups of about fifty students simultaneously. The instructions for the test are printed on one page, followed on the next page by the test itself. This enables the examiner to control the amount of time allotted to reading the instructions and the amount of time for the test proper. The Gordon Directions Test has the instructions printed on the

same page with the test itself since the test score is intended to reveal the subject's alertness in comprehending the instructions. The time limits for the tests are as follows:

1. Completion test. 10 minutes.
   Special scoring method for each response. From two to four points are allowed for each perfect sentence.
2. Robinson Range of Interest Test. 20 minutes.
   This is ample time for all subjects to finish the test. The test is really given by the amount limit method since every candidate is expected to finish the whole test. Scoring formula: $S=100-2W-O$ in which
   S = Total Score
   W = Number of wrong responses
   O = Number of omitted responses
3. Gordon Directions Test. Part 1, one minute; Part 2, one and one-half minutes; Part 3, two minutes; Part 4, two and one-half minutes. The score is the total number of right responses on the last three pages.
4. Analogies Test. Two minutes.
   Scoring formula: $S=100+R-3W$ with the same notation as in the Range of Interest Test.
5. Marble Statue Test. This test is read by the examiner and the subjects are asked to reproduce the paragraph after an interval of thirty minutes filled with other tests. The score is the number of ideas reproduced. Just what constitutes an idea is left to the interpretation of the scorer with the help of the paragraph dissected more or less arbitrarily into ideas.
6. Opposites Test. Two minutes.
   Scoring formula: $S=R-W$ with notation as before.

The Criterion

The value of a mental test is determined solely by its agreement with a criterion. Much has been said about scholarship as a criterion. If our task is to ascertain whether or not a candidate is intellectually able to do college work, and if the students' retention, promotion, dismissal, or graduation is determined by his scholarship grades, then there can be no other criterion than scholarship. That seems so obvious that there is really no room for debate about the question. To use any other criterion constitutes an indictment against the scholarship marks as a criterion for retention and promotion of college students. If our tests agree well with estimated intelligence but fail to agree with faculty action concerning the dismissal of poor students and the promotion of good students, then

our tests are of little use to the administrative officers of the college, unless they are willing to be guided by our statement of the student's intelligence rather than their own scholarship ratings.

In view of what has been said *pro* and *con* regarding scholarship as a criterion for the efficiency of mental tests we have used both criteria. Our present series of six tests have been evaluated against the following criteria:

(1) Can we demonstrate that we can reduce the number of students who are dropped for poor scholarship or placed on probation for poor scholarship by the use of our mental tests for admission?

(2) How do our mental test ratings of all the students compare with the faculty opinion about the general ability of the students?

The first criterion refers only to those who were pronounced as failures and dropped from college for inability to do college work or placed on probation as doubtful students with two thirds of the regular program. The second criterion has reference to the whole class including the good students.

The second criterion by which our tests have been evaluated is the median instructor's estimate of the student's general ability. We sent the following letter to all the instructors who have freshmen in their classes:

> *To the Members of the Faculty:*
>
> In order to evaluate the psychological tests which were given to the freshmen recently it is necessary to ask for your co-operation in judging the students. Those tests which agree fairly well with the pooled judgment of the faculty will be retained and the tests which fail in this regard will either be improved or cancelled.
>
> Kindly place a check mark in one of the ten spaces after each name to indicate the student's GENERAL ABILITY compared with the FRESHMAN CLASS. General ability is to be thought of as broader than mere scholarship. Leave unmarked the names of students you do not know or whom you have not had in your class for at least one semester. Try to place approximately the same number of students in each of the ten classes.
>
> Please return this report to the Dean's office by ———.

Attached to the letter was a list of the names of the Freshmen with ten numbered spaces after each name. When the returns were complete the median instructor's estimate was determined for each student and this was used as a criterion for the tests. These instructors' judgments were given in March, 1918. The tests were given in October, 1917.

Method of Evaluating the Mental Tests

There are two fundamentally distinct problems connected with the evaluation of a mental test. These are (1) Is the test diagnostically significant when applied to each individual student? and (2) What is the degree of relationship between the two variables without reference to any individual subject? These problems are restatements of the two problems in the preceding section.

The Pearson coefficient of correlation and the correlation ratio express the degree of relationship between the variables as such and the regression equation can be used for predicting one of the variables when the other is known. By means of the standard error of estimates we may quantify the reliability of the prediction for any individual but this evidence is not as satisfying as when the results are stated in terms that bring out the individual members of the group. The combination of tests can be handled by multiple correlation but the procedure is in practice hardly worth the trouble if our problem centers on the diagnostic value of the test for the individual rather than for the group. However, in order to extract all possible information from our tests they have been analyzed both by correlation methods referring to the group as a whole, and by inspection of the scatter diagrams referring to the individual student. I shall refer to the latter procedure as *the method of critical scores.*

The Method of Criterial Scores

The median instructors' estimate is ascertained for each student. This is based on the opinion of from three to six instructors. A scatter diagram is plotted for each test. Figure 1 is the scatter diagram for the Trabue Completion Test. The Pearson coefficient of correlation between the pooled estimate of the instructors and the score in this test is +.49.

The next step is to determine an upper critical score and a lower critical score. These are indicated in figure 1. They are so drawn that every student below the lower critical score is below the mean of the instructors' estimate. Every student above the upper critical score is above the average in the opinion of the faculty. Some tests have fair correlation coefficients but do not yield any critical scores on account of exceptions at the extremes. Some mental tests yield only one critical score. Thus when a test gives an upper critical score but no lower critical score we conclude that the test is useful for selecting unusually good students but not for spotting poor students.

The reader will notice that the lower critical score for the Trabue Completion Test is at 48 and that all the nine students who scored below 48 were rated below the average by the faculty. Facts of this type are practically useful, and when combined with similar facts from other tests, they are diagnostically significant in qualifying and disqualifying students for college work.

Notice also that the upper critical score for the Trabue Completion Test (Figure 1) is at 84 and that all the students who scored above 84 in this test are rated above the average by the faculty. The horizontal double line on the chart indicates the mean of the faculty estimates.

When the critical scores have been ascertained for all tests, they are combined in order to enhance the reliability of the prediction. This is done as follows: On the student's mental test record card each score that is above its upper critical score is marked in red, and each test score that is below its lower critical score is marked in blue. The median percentile rank in all tests is also recorded. Then we obtain a list of all the students who have been dropped during the year for poor scholarship and also those who have been placed on probation or two-thirds schedule on account of poor scholarship. Our motive is to obtain a method of predicting these failures by our test records. In order to do this we prepare lists of students for conditions such as the following:

(1) A list of all students whose median percentile rank in six tests is 20% or below, 15% or below, 25% or below, etc.

(2) A list of all students whose median percentile rank in six tests is 20% or below, and who were below the lower critical score in two or more of the tests, three or more of the tests, one or more of the tests, etc.

(3) A list of all students whose median percentile rank in the six tests is 20% or below, and who were below the lower critical score in more tests than they were above the upper critical score.

The lists so provided are compared with the list of failures due to poor scholarship. We select as our standard method that which predicts the largest proportion of failures *without excluding any students who have made good.* This method intensifies the responsibility of the psychologist with regard to the individual. The correlation methods do not place this emphasis on the individual but state the abstract relation between the several variables and this does not bring home to the psychologist his chances of error in advising about an individual student.

After an extensive analysis of our test records and the list of failures we found that the following criterion was quite satisfactory:

Those students are reported as doubtful who attain a median percentile rank of 22% or below, and who are below the lower critical score in two or more of the six tests.

The Mental Test Rating

In order to quantify this test analysis into a scale we have adopted the procedure of designating as the mental test rating the median percentile rank in all six tests plus five points for each test in which the student is above the upper critical score and deducting five points for each test in which he is below the lower critical score. We report as doubtful all students whose mental test rating is +10 or below.

By inspecting Figure 2 the reader will notice that we could have eliminated seven out of the eleven total failures at the beginning of the year. Instead of eleven total failures there would have been only four if the mental test rating had been a criterion of college entrance. Furthermore, we should have eliminated at the beginning of the year eight of the seventeen students who were placed on probation for poor scholarship. Instead of seventeen such cases there would have been only nine cases of probation for poor scholarship. Another very important fact is that not one of the students who were below the lower critical mental test rating was acceptable as a student. All of them without exception were either dropped for poor scholarship or placed on probation with reduced program on account of poor scholarship. They should have been saved from the discouragement which comes with failure and should have been advised to take up some other work. None of the acceptable students scored below the lower critical mental test rating.

Figure 2 also shows that the eleven students who were rated as unusually good students by the faculty were all above the average according to the mental test rating.

In figure 3 we have the scatter diagram for instructors' estimate against the combined mental test rating.

The Pearson correlation coefficient for these two variables is +.60. It is interesting to know that the application of the method of multiple correlation to these two variables for a closely parallel set of six tests gives a total coefficient, R, of +.67 which is not alarmingly superior to the method of critical scores and certainly

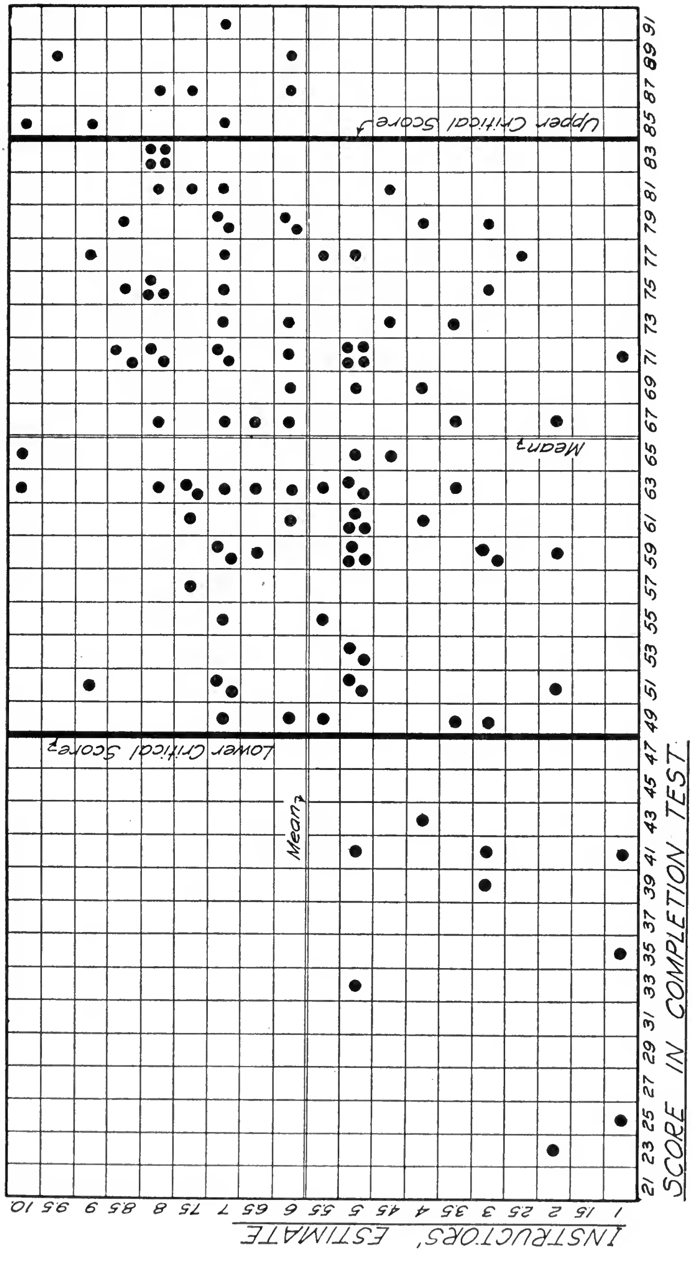


FIGURE 1.

Scatter diagram showing relation between median instructors' estimate and score in completion test. After the means have been drawn the two critical scores are so selected that all students above the upper critical score are above the average in the opinion of the faculty and all students below the lower critical score are below the average in the opinion of the faculty.

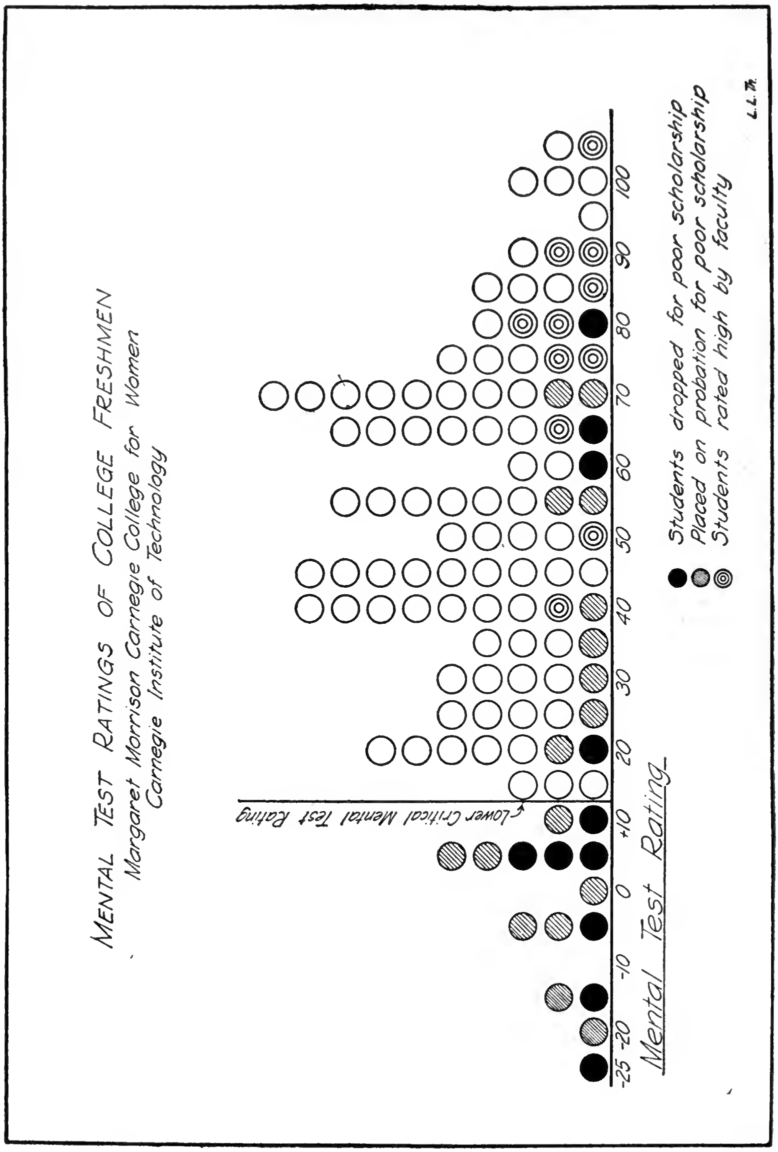


FIGURE 2.

Frequency distribution of mental test rating which is the median percentile rank in six mental tests adjusted by adding five points for each test in which the candidate is above the upper critical score, and deducting five points for each test in which the candidate is below the lower critical score. Note that seven of the eleven freshman total failures were below the lower critical mental test rating.

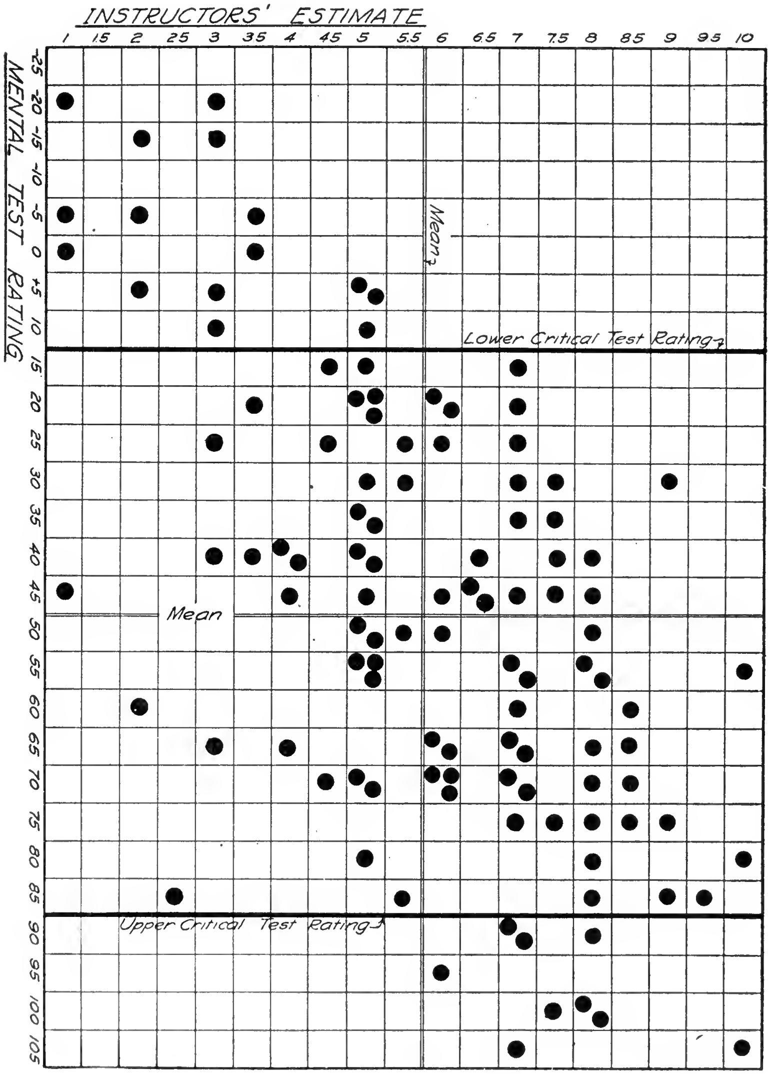


FIGURE 3.

Scatter diagram showing relation between instructors' estimate and menta ltest rating. Note that all the freshmen rated below the lower critical mental test rating (+10.) are below the averagein the opinion of the faculty and that all who scored above the upper critical mental test rating (85) are rated above the average in the opinion of the faculty.

FIGURE 4.
Frequency distribution of instructors' estimate showing the same facts as in figure 3.

not nearly as controllable. Figure 3 shows that all students below the lower critical mental test rating were rated below the average by the faculty and that all students above the upper critical mental test rating were rated above the average by the faculty. In figure 4 we have these last two statements isolated from the scatter diagram of Figure 3.

Summary and Recommendations

Mental tests have been demonstrated to constitute a useful criterion for admission to college. We shall have five criteria for entrance, namely (1) high school credits, (2) high school principal's estimate of the applicant's capacity, (3) college entrance examinations, (4) Dean's interview with the applicant, and (5) psychological tests. We have analyzed the relative efficiency of the five criteria and have found that mental tests which occupy one hour predict freshman scholarship more accurately than the second semester's scholarship can be predicted from that of the first semester.

The mental test rating would have eliminated seven of the eleven total failures in the particular freshman class described. If the mental tests had been a criterion for admission there would have been only four failures instead of eleven.

No average or good student would have been eliminated by the mental test rating. All students who scored below the lower critical mental test rating were, without exception, poor students. All of them were either dropped from college entirely in the first year or placed on probation with two-thirds schedule on account of poor scholarship. Hence no individual injustice would result from the use of mental tests as a criterion of college entrance.

All of the freshmen who were rated high by the faculty were above the average in the mental test rating.

The standard series of six tests has been given to the freshman class of 1918–19 but in addition four new tests have been given which are in the process of being accepted or rejected. If any of the new tests prove to be more successful than last year's tests they will be added to the standard series and the old tests discarded. By continuing this process the efficiency of the test series as a criterion for college entrance will probably be raised every year.

The method of critical scores has been applied successfully in standardizing the Army Trade Tests. The method is more controllable than the method of multiple correlation for establishing diagnostically significant criteria. One of its main advantages is

in the ease of application. It emphasizes prognosis with respect to each individual student rather than the abstract statement of relationship between the variables concerned.

The method of critical scores consists simply in plotting a scatter diagram for each test, showing relation between scholastic attainment and mental test score for each individual. An upper critical score is selected such that all who score above it are above the average in scholastic attainment. A lower critical score is selected such that all who score below it are below the average in scholastic attainment. Some mental tests do not yield any critical scores.

When tests are combined by the method of critical scores the median percentile rank is corrected by adding five points for each test in which the subject is above the upper critical score and deducting five points for each test in which the subject is below the lower critical score.

This emphasizes the extremes. The same effect of emphasizing the extremes is obtained by using the standard deviation but it is not so readily interpreted. I do not wish to discourage the use of multiple correlation as a means of analysis but I do insist that we should not be enslaved by its laborious refinements in making mental test results practically applicable.

THE NECESSITY OF TEACHING DERIVED FORMS IN SPELLING

E. HORN AND E. J. ASHBAUGH

University of Iowa

Introduction. Progress in the scientific study of spelling has been made along three major lines: first, in selecting and grading the words which are to make up the course of study; second, in the experimental determination of the most economical methods of learning; and, third, in the development of satisfactory scales for the measuring of spelling ability. In all of these lines research has proceeded well beyond the pioneer stage so that we may now expect a series of studies which will greatly refine our present knowledge and practice. The investigation described in this article was undertaken in the attempt to accomplish one such refinement.

From the beginning of vocabulary studies in spelling, the writers believed that each change in the form of a word should be scored as a separate word, and that attempts to group according to root forms were objectionable. Particularly objectionable is the practice of deciding, without data, whether or not a given form constitutes special spelling difficulty. The following statement from Ayres, *Measuring Ability in Spelling,* illustrates the use of judgment in determining whether or not a given form is difficult to spell.

"In making up this list, there has been no attempt to reduce all the words to a dictionary basis. Instead the attempt has been to include all the forms of the words which present different spelling difficulties. Thus the various forms of the verb "be" are included as separate words because they present separate spelling difficulties. In the same way, "man" and "men," "women and "women," are included for the same reason. On the other hand, plurals and verb forms presenting no characteristic spelling difficulties beyond those inherent in the singular or infinitive have not been included. This procedure has necessitated making many arbitrary decisions, but in each case the controlling purpose has been to make each decision on the basis of spelling difficulty."

The data which follow show conclusively that in the great majority of cases a change in form is accompanied by a difference in spelling difficulty, and that with few exceptions derived forms are more difficult than the roots upon which they are built. According-

ly it becomes important to know the spelling difficulty of each derivation and also the frequency with which it occurs in correspondence.

All data in Tables I, III, IV, V, and VI are based upon the spelling of a random selection of two hundred children from the second grades in the schools of Iowa. The words chosen for Table I were those root forms upon which the children spelled with an accuracy of 60 percent. or better, and for which the accuracy on one or more derived forms was also available. The words in Tables III, IV, V, and VI were chosen by taking in regular order from the alphabetized list words appearing in both root form and with the suffix under consideration.

Table I shows the accuracy of spelling of second grade children upon the root and derived forms of forty words. The table is read as follows: The accuracy of spelling of second grade children on the word *be* was 89 percent. and on the derived form, *being*, was 65 percent.,—a difference of 24 percent. In all cases in this and the following tables, when the accuracy on the derived form was less than on the root form, the difference is marked minus; when the accuracy on the derived form was greater than on the root form, it is marked plus.

The seventh word has two derived forms and is read thus: The accuracy of second grade children on the word *look* was 82 percent., on *looking* was 70 percent., and on *looked* 40 percent. The difference between the accuracy on the root form and the first derived form was 12 percent. and between the root and second derived form was 42 percent. The accuracies and differences on all other words in the table are read in the same manner.

The last column at the right is headed "Group Average." Under this is shown for the first and second groups of twenty words separately the average difference between the accuracy on the root form and on the first, second and third derived forms. It will be noted that the average difference in accuracy between the root form and the derived form is greater in the group with the higher accuracy on the root form.

At the foot of the table is given the averages and medians for accuracies and differences on the entire list of forty words. The decrease in accuracy of spelling on the derived forms is really greater than appears in the average differences since they must be thought of in terms of the accuracy on the root forms upon which they are

TABLE I

Difference between root and derived form in accuracy of spelling.

root	root acc.	Accuracy of derived forms 1st	2nd	3rd	Difference from root Acc. 1st	2nd	3rd	Group Ave.
be	89	ing 65			−24			
go	89	ing 68			−21			
see	88	ing 71			−17			
lone	84	ly 14			−70			
come	83	ing 46			−37			
do	83	ing 72			−11			
look	82	ing 70	ed 40		−12	−42		1st
tell	82	ing 64			−18			−32.65
have	81	ing 39			−42			
will	81	ing 44			−37			
big	80	ger 23	gest 5		−57	−75		
land	80	ing 58			−22			2nd
old	80	er 42	est 36		−38	−44		−51.8
say	80	ing 52			−28			
hand	79	ed 43			−36			
run	77	ning 9			−68			
dear	76	ly 29	est 29		−47	−47		
play	76	ing 68	ed 25		− 8	−51		
cold	73	er 42			−31			
think	73	ing 44			−29			
bad	70	ly 37			−33			
love	69	d 40	ly 40	ing 15	−29	−29	−54	
box	68	es 34			−34			
like	68	ly 32	d 17		−36	−51		
call	67	ing 45	ed 33		−22	−34		
put	66	ting 11			−55			
fast	65	er 49			−16			
fill	65	ing 34	ed 21		−31	−44		
find	65	ing 54			−11			
glad	64	ly 32			−32			1st
name	63	d 29	ly 22		−34	−41		−26.55
bill	62	ed 14			−48			
kiss	61	es 24			−37			
part	61	ly 24			−37			2nd
send	61	ing 35			−26			−34.56
want	61	ing 49	ed 26		−12	−35		
hold	60	ing 55	er 33		− 5	−27		
live	60	d 46	s 43	ing 22	−14	−17	−38	3rd
stand	60	ing 60			0			−45.3
wish	60	ing 41	es 27	ed 16	−19	−33	−44	
Ave.	72.3	42.7	28.4	17.7	−29.6	−40.7	−45.3	
Median	71.5	43.0	28.0	16.0	−30.0	−41.5	−44.	

SPELLING

TABLE II

Difference between root and derived form in frequency of occurence. Grade II

Frequency of derived forms						
root		1st		2nd	3rd	4th
be	4042	ing	305			
go	656	ing	640			
see	824	ing	14			
lone	16	ly	12			
come	882	ing	312			
do	1309	ing	235			
look	320	ing	109	ed 50		
tell	359	ing	20			
have	4781	ing	331			
will	4211	ing	68			
big	278	ger	19	gest 15		
land	249	ing	13			
old	215	er	71	est 8		
say	666	ing	43			
hand	308	ed	31			
run	111	ning	32			
dear	2194	ly	7	est 21		
play	84	ing	15	ed 23		
cold	220	er	5			
think	650	ing	54			
love	282	d	8	ly 51	ing 19	
box	120	es	24			
like	744	ly	15	ed 33		
call	334	ing	50	ed 123		
put	243	ting	40			
fast	28	er	8			
fill	102	ing	26	ed 40		
find	510	ing	17			
glad	760	ly	42			
name	423	ly	9	d 15		
bill	170	ed	6			
kiss	8	es	13			
part	217	ly	12			
send	657	ing	194			
want	583	ing	19	ed 118		
hold	157	ing	28	er 25		
live	113	d	15	s 21	ing 47	
stand	68	ing	20			
wish	539	ing	43	es 205	ed 66	

based. For example, the average difference of the first derived forms is 29.6 percent. Since the average accuracy on the root forms was 72.3 percent. this 29.6 percent. represents a decrease of 40.9 percent. The average difference of the second derived forms is 40.7. But since only fourteen words are represented by second derived forms, this must be considered in terms of the accuracy of these fourteen words or 69.1 percent. and hence represents a decrease of 58.9 percent. In like manner the average of the three third derivatives represents a decrease of 71.9 percent.

Table II gives for each root and derived form in Table I the frequency of occurance in two studies in which the words have been scored separately. One of these studies,[1] represents an analysis of 1175 letters of bankers to bankers; the other,[2] an analysis of 3763 letters written by Iowa correspondents. These two studies represent the scoring of over 400,000 running words.

The tables should be read: *be,* has a frequency in these combined lists of 4042; *being,* a frequency of 305, etc. The only case in which the derivation occurs more frequently than the root form is in the word kiss, which is exceeded in frequency by kisses.

Table III represents the difference between the accuracy of second grade children in spelling root forms and the same words with the suffix *ing.* The first twenty words attach the suffix without any change in the form of the root word and the second twenty drop the final *e* before adding *ing.* The table is read as follows: The accuracy of second grade children on the word *ask* was 54 percent. and on the derived form, *asking* was 42 percent. Hence the difference between the two forms is 12 percent. The percentages on the other words are read in the same manner. Differences showing the derived form less than the root are proceded by a minus sign, and those showing the derived form greater, by a plus sign.

The table has been divided horizontally to indicate groups of words upon which the accuracy on the root forms fell within the same quartile percentages, that is 100 percent.—76 percent.; 75 percent.;—51 percent.; 50 percent.—26 percent, and 25 percent.—

---

[1] Ernest Horn. *A Study of the Vocabulary of Letters Written by Bankers to Bankers about Banking.* Study unpublished.

[2] W. N. Anderson. *The Determination of a Spelling Vocabulary Based upon Written Correspondence.* Ph.D. Thesis, University of Iowa, 1917.

TABLE III

Difference between root and derived form. Grade II.

Word	Freq. Root	Freq. Der.	Acc. Root	Acc. Deriv.	Dif. in Acc.	Aver. dif. in group
ing–						
ask	263	67	54	42	−12	
bank	569	299	53	30	−23	
belong	17	7	53	32	−21	
bring	128	19	59	49	−10	
call	334	50	67	45	−22	−18
act	56	7	31	17	−14	
add	77	6	28	32	+ 4	
amount	272	33	29	5	−24	
buy	3	51	40	20	−20	
check	224	31	33	12	−21	
clean	55	19	35	27	− 8	−14
answer	250	123	10	10	0	
attend	74	19	20	28	+ 8	
await	13	69	16	15	− 1	
board	129	6	22	9	−13	
break	21	9	7	12	+ 5	
build	41	110	12	15	+ 3	
carry	86	33	23	27	+ 4	
coast	17	5	10	12	+ 2	
collect	36	9	6	0	− 6	+0.2
come	882	312	83	46	−37	
have	4781	331	81	39	−42	−40
bake	15	6	48	31	−17	
change	298	16	30	14	−16	
inclose	15	12	38	25	−13	
leave	143	31	39	10	−29	−19
advertise	18	50	8	9	+ 1	
advise	242	34	16	10	− 6	
anticipate	13	6	0	2	+ 2	
appreciate	308	10	1	1	0	
arrive	37	7	4	1	− 3	
assure	94	70	2	4	+ 2	
believe	295	11	8	6	− 2	
choose	22	8	19	14	− 5	
dance	21	11	20	16	− 4	
debate		5	20	17	− 3	
desire	192	7	12	15	+ 3	
figure	59	11	8	3	− 5	
increase	54	28	4	7	+ 3	
invite	1	29	18	23	+ 5	− 0.9

TABLE IV

Difference between root and derived form. Grade II.

	Word	Freq. root	Freq. Der.	Acc. Root	Acc. Deriv.	Dif. in Acc.	Aver- dif. in group
-ed							
	ask	263	122	54	31	−23	
	bill	170	6	62	14	−48	
	call	334	123	67	33	−34	−35
	act	56	10	31	7	−24	
	add	77	44	28	42	+14	
	check	224	6	33	7	−26	
	clean	55	29	35	14	−21	
	cover	121	31	35	14	−21	−16
	accomplish	14	6	3	1	− 2	
	address	224	100	19	8	−11	
	adjust	12	14	22	7	−15	
	appear	36	18	6	1	− 5	
	attach	10	22	9	9	0	
	attend	74	13	20	3	−17	
	allow	53	26	10	6	− 4	
	answer	250	40	10	12	+ 2	
	connect	19	21	1	2	+ 1	
	consider	164	49	7	5	− 2	
	contain	30	12	9	12	+ 3	
	correct	54	7	10	1	− 9	− 5
-d	agree	421	8	32	20	−12	
	care	283	21	40	27	−13	
	change	298	45	30	14	−16	
	charge	230	1	30	13	−13	−13.5
	advance	88	45	9	1	− 8	
	advertise	18	15	8	5	− 3	
	advise	242	26	16	4	−12	
	appreciate	308	52	1	3	+ 2	
	arrange	67	21	3	3	0	
	arrive	37	48	4	4	0	
	associate	5	7	0	1	+ 1	
	assure	94	25	2	2	0	
	cause	57	22	11	11	0	
	continue	70	32	9	8	− 1	
	convince	15	9	9	5	− 4	
	decide	91	82	6	6	0	
	deserve	8	6	5	2	− 3	
	desire	192	38	12	7	− 5	
	dispose	9	11	18	22	+ 4	
	figure	59	11	8	6	− 2	− 2

TABLE V

Difference between root and derived form of words adding er. Grade II

Word	Freq. Root	Freq. Der.	Acc. Root	Acc. Deriv.	Dif. in Acc.	Aver. dif. in group
old	215	71	80	42	−38	
bank	569	25	53	45	− 8	
cold	220	5	73	42	−31	
cool	28	7	52	35	−17	
farm	186	92	54	36	−18	
feed	27	9	54	39	−15	
hold	157	25	60	33	−27	
kind	507	6	51	35	−16	
long	374	60	54	40	−14	
low	71	28	52	16	−36	
near	182	9	52	31	−21	−20.3
clean	55	10	35	27	− 8	
high	280	37	26	24	− 2	
own	243	63	36	5	−31	
paint	57	8	27	15	−12	
print	23	7	50	29	−21	
rent	63	9	46	33	−13	−14.5
build	41	7	12	8	− 4	
cheap	31	12	10	9	− 1	
commission	76	7	2	2	0	− 1.7

0 percent. The average differences within these groups are shown to the right and the same fact is noted as in Table I, namely, that the average differences are greater in groups having the higher accuracy upon the root forms.

Table IV presents the same kind of data upon forty words with the suffix *ed* as is given in Table III. Table V presents twenty words with the suffix *er* and Table VI presents twenty words with the suffix *ly*. The table is made up in exactly the same manner as Table III, is read in the same way and the results show the same thing.

From the data given above, the authors conclude, first, that in selecting words to make up the course of study in spelling each change in spelling should be incorporated as a separate word; second, that in teaching these words each separate form should be taught as a different word; third, that in making a scale for the measurement of spelling ability the spelling difficulty of each word must be discovered and the word assigned to its appropriate place in the scale.

TABLE VI

Difference between root and derived form of words adding ly. Grade II.

dear	2194	7	76	29	−47	
deep	16	9	53	23	−30	
glad	760	42	64	32	−32	−31
clear	75	10	44	13	−31	
fair	142	34	46	13	−33	−32
absolute		47	4	8	+ 4	
according	3	4	7	2	− 5	
actual	21	20	1	1	0	
annual	65	14	2	2	0	
awful	99	37	10	3	− 7	
certain	72	223	4	3	− 1	
complete	60	12	6	5	− 1	
correct	54	8	10	2	− 8	
direct	95	22	9	2	− 7	
entire	60	96	7	11	+ 4	
exact	22	24	0	4	+ 4	
extreme	8	13	5	0	− 5	
final	23	20	15	2	−13	
friend	485	12	16	19	+ 3	
general	140	16	5	0	− 5	− 2.5

COMPARATIVE TESTS OF HOME WORK AND SCHOOL WORK

W. H. HECK*

University of Virginia

The following tests were given on Tuesday and Wednesday, December 1 and 2, 1914, in the grammar and high school grades (5A–12) of the public school at Harrisonburg, Virginia. This town, with a population at that time of nearly 5,000 had long had a high average of comfortable homes and a general interest in public education. The students taking the test were representative of all classes in the community. The weather was good during the two days and the school work normal, the time having been selected so as to avoid inside or outside interference.

Grammar Grade Tests in Arithmetic

The test material was composed of four separate tests: (1 and 2) modifications of the very similar first and second forms of the Courtis test No. 7 in computation, and (3 and 4) the very similar first and third forms of the Courtis test No. 8 in reasoning problems. These tests represented practical arithmetic lessons, not artificial combinations, and the author had found them very satisfactory in his experiments on mental fatigue. The computation tests consisted of three examples in each of the four fundamental operations, and the reasoning tests consisted of eight problems using various combinations of the four fundamental operations. No fractions of any kind were involved, unless through a student's mistake. The time for any test need not have been as much as 15 minutes.

The total value of the examples in each of the computation tests was 16 (the individual values being ½, 1, or 2, with corresponding deduction for mistakes); the total value of the eight problems in each of the reasoning tests was eight (1 each); therefore, each student could have received a total of 48 points in the four tests (24 in his school work and 24 in his home work).

The half-year grades tested were 5A, 6B, 6A, 7B, 7A, and 8, the average age of the students by classes being approximately 12–15 years. These 6 grades were divided into two groups, as equally balanced as possible; the first group was composed of grades 8, 7B, and 6B, and the second group of grades 7A, 6A, and 5A. Only the class teachers were present with the children during the author's explanation, assignment and collection of each of the four test papers.

*Shortly after the completion of this article the author succumbed to an attack of pneumonia.—Ed.

There was no supervision of the student's work on the tests, and the teachers went on as usual with their regular instruction and discipline. Special supervision of this work by the class teacher would not have been typical of most study periods in school and would have given the school work undue advantage over the home work in the present experiment.

As rapidly as practicable after the opening of school on Tuesday, test 1 (computation 1) was assigned to the first group, and test 3 (reasoning 1) to the second group. The students were told that they could spend plenty of time in their study periods (when another division of the same class was reciting) to work out all the examples carefully; that the teachers would accept a reasonable excuse if the test hindered somewhat the preparation of the regular work; that the rating of the papers would not be done by anyone connected with the school and would not count against them individually; that the marks would be averaged by classes so as to show how well each class had profited by its arithmetic instruction; that any assistance of one student by another would be dishonest and would confuse the result. The students accepted the work gladly as a variation in the school routine and as a chance to show what they could do.

The papers were collected toward the close of the afternoon session, and test 3 (reasoning 1) was then assigned to the first group and test 1 (computation 1) to the second group, the work to be done by each student at home without help from anybody else and to be handed in after the opening of school next morning. Nothing at all was said about a comparison between home and school work, and the children did not show at any time during the series the slightest suspicion as to the final purpose of the tests. On Wednesday morning the six sets of papers were collected, and then test 4 (reasoning 2) was assigned to the first group and test 2 (computation 2) to the second group. These papers were collected toward the close of school, and test 2 (computation 2) was assigned to the first group and test 4 (reasoning 2) to the second group. These last papers were collected Thursday morning.

The first group did the first computation and the second reasoning test in school on Tuesday and Wedneday, respectively, while the second group was doing the first reasoning and the second computation test; and the first group did the first reasoning and the sec-

ond computation test at home on Tuesday and Wednesday, respectively, while the second group was doing the first computation and the second reasoning test.

The practice effect was thus so divided between the school work and the home work as to be neutralized and consequently in favor of neither. It was Schmidt's[1] failure to give due weight to the pratice effect in his methods and results, that furnished the initial suggestion to the author for the present experiments. The practice effect here, however, is slightly negative in the easy computation work (probably on account of increasing carelessness) and but slightly positive in the case of the more difficult reasoning problems. If we leave out of consideration for the moment or neutralize by a rearrangement of the marks, any possible difference between home and school work, we find that the average mark of all the students in the second computation test is 2.7 percent. less than that for the first, and the mark for the second reasoning test is 4.9 percent. more than that for the first.

Taking the 141 students or the six classes as a whole, the school work with its one computation and one reasoning test was as nearly equal as practicable to the home work with its one computation and one reasoning test, the only general difference being that between the school environment and the home environment. And a measurement of the influence of this difference in environment upon the work of school children was, of course, the main object of the present experiments.

The results were as follows:

TABLE I

Average Mark

Class	Students	School Work	Home Work	Difference
8	24	23.00	22.79	−0.21
7A	21	20.62	19.62	−1.00
7B	22	20.50	20.86	+0.36
6A	20	20.75	21.45	+0.70
6B	23	20.43	18.22	−2.21
5A	31	17.52	19.65	+2.13
Total	141	20.25	20.39	+0.14
Average of Class Averages	23.5	20.47	20.43	−0.04

[1]Archiv fur die gesammte Psychologie, 1904, vol. III, pp. 33–152; Zeitschrift fur Experimentelle Paedagogik, 1907, vol. IV., pp. 189–210, vol. V, pp. 23–50. These articles and the somewhat similar report by Meumann can be found in Heck's *International Source Book on Home Study*, to be published as a bulletin of the U. S. Bureau of Education.

It is easily seen from Table I that the difference is so close to 0 as to be moved a tiny fraction above or below by the method of calculation: 0.14 point, or seven-tenths of one per cent., in favor of home work in case of an average of all the students, and 0.04 point, or two-tenths of one per cent., in favor of school work in case of an average of class averages. Three of the grades (2 in the first group and 1 in the second) are slightly in favor of school work, and three (1 in the first group and 2 in the second group) are slightly in favor of home work—about as equal a balance as possible. It does seem that the two lowest classes, 6B and 5A, show preferences which are more than a negligible chance variation; but these perferences are almost equally balanced against each other.

Home work in arithmetic can be done as well as school work, if the children will to make it so; and objections on account of its supposed inferiority to school work do not seem to be justified. As a matter of daily routine the children may not know how or may not will to do either well. But it is precisely the function of good teaching, and of the home in co-operation, to meet such needs.

High School Tests in English Composition

On Tuesday morning the author assigned to the first high school group (consisting of 30 students from the first year and 11 from the fourth year class) a composition of between 500 and 1,000 words on the subject, "The Harrisonburg High School." The papers were to be collected just before the close of the afternoon session. The students were told to take as much time as they needed to do their best, although preparation for or even attendance at some class period might have to be sacrificed. Nothing was said as to making one or more copies, this being a matter for the student's own judgment and habit. It was explained that the marks on the papers were not to count for or against the students individually but were to show how well the different classes were profiting by their English instruction. These compositions were written during the usual study periods in the assembly hall and were not specially supervised by the principal in charge of the hall. After the collection of the papers a second composition, "The Town of Harrisonburg," was assigned to the first group, to be written independently and honestly by each student at home and handed in at the opening of school the next morning. In a similar way the first composition, "The Harrisonburg High School," was assigned as home work to

the second group (consisting of 21 students of the second year and 13 of the third year class). The two sets of papers were collected on Wednesday morning and the second composition was then assigned as school work to the second group. The few compositions which were handed in later, because they had to be finished or copied in school, were put together separately but without any comment or sign being made of the fact. Not only were these papers thrown out but so were the past or future test compositions by the same students, and also those by any student who wrote only one. On Thursday the students were surprised when told that the main purpose of the tests was a comparison between home work and school work.

It must be noted that the subjects for the two compositions were such that any student of high school grade ought to have had enough facts in mind to write a composition easily; certainly there was no need to consult outsiders or literature, a need that often brings injustice to some students who have inferior opportunities for such assistance. Not the getting of facts but the selection from the many at hand, and then the individual expression of ideas in regard to them—these were the real requirements and the two compositions were very nearly alike in their demands as to content and as to form.

The first composition was written in school by the first group and at home by the second group; the second composition was written at home by the first group and in school by the second group. This balancing was done primarily to neutralize the practice effect. However, the practice effect was really negligible in the present experiment, as the second composition was 0.2 per cent. less in form and 1.6 per cent. more in content than the first.

Although the grammar grade tests in arithmetic were easily marked and compared by the author, the case was far different and more difficult with the high school tests in English composition. In fact, these 150 papers were put aside on account of the pressure of other work until the last quarter of 1917, when Mr. W. C. Whitlock, editor of the students' magazine at the University of Virginia, was employed to mark the papers, so that a report on the tests could be included in the present report. After rejecting as futile the attempts to apply some so-called standard system of grading to these papers, the author decided on a double plan: one mark for

form—based on deductions for the mistakes in spelling, grammar, punctuation, handwriting, etc.; and one mark for content—the selection, organization, and interpretation of facts. (Both the formal and the material errors emphasized by Schmidt correspond to those included in the first mark here, as he neglected the more important and positive values of content.)

The marking for form was simple. Each mistake counted one and the total number was divided by the total number of words in the composition; then the resulting percentage was subtracted from 100, the remainder being the mark for form. Of course, this method has its deficiencies: some mistakes have little relation to words and the number of them; some words (the articles, for instance) are free from mistakes and are very frequently used; there may be more than one mistake in the same word; repetitions of the same mistakes are not of the same value as new mistakes; different mistakes have different practical values. Nevertheless, this method gives consideration to the necessity of proportioning mistakes to the length of the composition, as it would be manifestly unfair to compare *per se* the mistakes of form in one composition with those in another of twice the length. However, any class method that is exactly and impartially applied to all the compositions would serve the present purpose of comparing school and home work. As the average number of mistakes in each group is given in the table, other comparisons of these mistakes are made possible, if desired.

In regard to content, the best method of approach to exactness and impartiality in marking the papers seemed to be the reliance upon the judgment of one person as to the comparative merits of each and all of the papers. One trained mind reacting to 150 compositions on two clear and closely allied subjects gave more promise than any system planned to be used by many minds on many types and subjects of compositions. But the attainment of impartiality in this single judgment of content required several strict conditions, which were conscientiously carried out in the present case. (1) The mark for form was not to influence the mark for content, except in so far as the content was directly affected by the mistakes in form. (2) The marking was to be done by a person entirely ignorant of the grouping, order,etc., of the papers for comparison of home with school work. The papers were, therefore, shuffled into a confusing disarray before they were given out to be

marked. (3) The marking of each paper was to be done without any recollection of or reference to the previous content mark of the other paper by the same student. (4) The marking of each paper was to be done by a person fresh in mind, so that fatigue or boredom would not warp the judgment. Only a few papers were gone over at one time, and the marking of the 150 papers was extended over three months.

The results were as follows:

Class	Students	Number of Words		Number of Mistakes		Form Mark		Content Mark	
		School	Home	School	Home	School	Home	School	Home
Total	75	499	546	44	45	91.18	91.57	80.59	81.30
First Group	41	465	528	39	42	91.61	92.05	79.51	81.15
Second Group	34	541	567	51	49	90.57	91.36	81.88	81.62
Boys	31	476	550	51	53	89.29	90.36	78.81	80.74
Girls....	44	515	543	39	39	92.43	92.82	81.84	81.68
Reduced Total	64	492	534	45	45	90.85	91.57	80.53	81.16

Table II shows that the home compositions as a whole are but 0.4 per cent. better in form than the school compositions and but 0.9 per cent. better in content—differences so small as to be negligible in school practice. To be sure, the very small percentages in favor of the home compositions and the additional fact that these compositions are 9.4 percent. longer than the school compositions might imply slightly more freedom and more thoughtfulness of written expression in the home environment; but this suggestion has very little proof. (Schmidt believes that greater length of the compositions implies less thoughtfulness.)

Furthermore, the lack of any important difference between the home and the school compositions is as noticeable when the smaller groups are compared with each other. With a possible practice effect in their favor, the home compositions in the first group are 0.5 and 2.1 percent. better in form and content, respectively, than the school compositions; and with a possible practice effect the other way, the home compositions in the second group are 0.9 percent. better in form and 0.3 percent. worse in content. In

order to see whether the difference was greater or less in case of the boys as compared with the girls, averages were made for the 31 boys (15 in the first group and 16 in the second) and for the 44 girls (26 in the first group and 18 in the second). The home compositions of the boys were found to be 1.2 percent. and 2.4 percent. better in form and content, respectively, than their school compositions, and the home compositions of the girls were found to be 0.4 percent. better in form and 0.2 percent. worse in content than their school compositions. The smaller differences in the case of the girls are due to the proportionately better work of the girls in school. In order to make a total with the first and second groups equal in number of boys and of girls represented and in number of students in the classes balanced against each other, the first papers alphabetically in the various sets for boys and for girls were eliminated until there were 15 boys and 17 girls in each group and 21 first year and 11 fourth year students in the first group and 21 second year and 11 third year students in the second group. In case of this equality in the two groups, the 64 home compositions were only 0.8 percent. better in both form and content than the 64 school compositions. There seems no way of getting from the results of these tests any difference one way or the other between home and school compositions, which is large enough to affect school practice.

A significant fact is revealed by the amount of individual variation in the home as compared with that in the school compositions. The standard deviations from the average marks of the home compositions for form and for content are 3.29 and 4.06, respectively; while the standard deviations for the school compositions are 3.83 and 6.05, respectively. The deviations in the school compositions are 16.4 percent. greater in the marks for form and 49.0 percent. greater in the marks for content than are those in the home compositions. The S. D. for the home compositions by the boys are 2.94 and 4.14 for form and for content, respectively; while the S. D. for their school compositions are 3.86 and 7.34, respectively. The S. D. for the home compositions by the girls are 2.59 and 3.95 respectively, and for their school compositions 3.27 and 4.44, respectively. In every case there is greater deviation in the school compositions, and more especially so in the less mechanical phase—that of content. And the boys, with greater variation throughout than is the case with the girls, show even proportionately more in their school, as compared with their home, compositions

This result is further corroborated by comparisons between the standard deviations of the home and school work in arithmetic. The 6 classes (141 students) had a S. D. of 2.25 in their home computation test and of 2.36 in their corresponding school test; and an S. D. of 1.42 in their home reasoning test and of 1.81 in their corresponding school test. These deviations are necessarily smaller than in the case of the compositions, because the possible range is much smaller, a perfect paper in computation getting a mark of only 16 and a perfect paper in reasoning problems a mark of only 8. The difference between the deviations in the marks on the reasoning problems are especially important, as these tests required more mental work than did the computation tests.

This greater variation in the work of students in the more or less uniform environment of the school, at least externally, than in the many homes with their multitude of differences, is cause for surprise. Evidently the influence of the school's uniform environment is very far from being uniform, and the supposed objections to home work on account of the varied influences at home apply more forcibly to the school work. The school has been and probably will long continue to be in part an artificial environment, to which children make adjustments with widely varying degrees of completeness and naturalness, and to which the individuality of their responses is suppressed in the daily routine of mass instruction and discipline. But a test, with its demands for individual work, brings out not only the differences in the children *per se* but the exaggerated differences of their reaction to the school environment in which the work is done. On the contrary, in the shelter of the home the child is more at ease, his lifelong adjustment to the persons and to the things there have given him the chance to express himself more normally, even in work for the school. It might be said that the greater variations in the school environment correspond more closely to the natural variations among children and that at home there is a general tendency to restrict and reduce these variations; but this explanation seems in itself unnatural. As nearly all tests of the variability in children's work have been made in school, it is probable that the educationists are tending to exaggerate the amount of real difference among children.

Assignments for home work requiring equipment, literature, assistance, etc., which many homes of the students concerned cannot furnish, are manifestly unjust and should not be given in public

schools. But for sensible assignments, calling for independent work on topics previously discussed in school, the conditions in the majority of homes are more natural and, on the whole, about as favorable for individual work as the conditions in the majority of schools, however much improvement is needed in both environments. The life-long adjustments of the children have tended to neutralize, even to utilize the differences in homes in expressing the more or less general spirit of family life and in meeting therein the duties of home and even of school. Anyway, these differences are far greater to the outsiders looking in than to the insiders looking out. And school people are generally outsiders looking in.

As a final conclusion from the experiments in arithmetic and in English composition, chosen to represent two main and somewhat contrasted types of school requirements, we may say that home work and school work are or can be made about equal to each other, are about equally dependent on good teaching, and will therefore develop as the teaching develops. There is no use in pitting one element in this trinity against another, when the problems of clear instruction and assignment, creation of interest, and strict checking-up of results involve each one mutually.

COMMUNICATIONS AND DISCUSSIONS

PSYCHOLOGY IN THE UNGRADED ROOM OF THE VILLAGE SCHOOL

The need for careful, scientific school management is, in all probability, as great in village schools as elsewhere, yet nowhere is the difficulty of securing it so great. For example, the ratio of low normal and sub-normal children to the total population of the school is fairly constant everywhere. The village school of six or seven hundred pupils needs, then, the services of a psychologist and a capable teacher of sub-normal children quite as much as does the city school. But, under present conditions, the village school *cannot* have a psychologist; and if it has the capable teacher of sub-normal children, such a teacher must be developed from material at hand. When an ungraded room is established under such conditions, how may the superintendent and the board of education have any assurance at all that the services rendered are performed in a reasonably efficient manner?

The services rendered by the organization of the ungraded room may be summed up in a general way as follows: (1) selection, (2) classification, (3) instruction. It is not the purpose of this paper to explain any of these functions in the organization and instruction of an ungraded class, but to suggest one possible check for the selection and classification of the pupils in the ungraded class, namely, the coefficient of correlation between the mental ages of pupils as found by the application of a general intelligence test and the scores made by pupils on Trabue Language-Completion Scales.

During May and June of last year and September of this year, regular classroom teachers recommended for examination about sixty children. Fifteen of these were selected, eleven of whom had attained the mental age of eight years or more as shown by the Yerkes-Bridges Point Scale. These ages were later corrected so as to read as of October first and used in the correlation table given below.

During the first week of October, the Trabue Language-Completion Scales, B and C, were used in measuring pupils of grades 2, 3, 4, 5, 6, 7, 8. Medians for each grade and attained age were found and represented on a graph. The graph was used as a norm in determining the Trabue scale mental age from the score of an individual.

The eleven pupils in the ungraded room who had attained the mental age of eight years or more were measured by the Trabue Language-Completion Scales, B and C, administered on two different days two weeks apart. Two pairs of scores were identical, six showed a difference of one point, two a difference of two points, and one a difference of four points. The aggregate score on the first test was 89, while the aggregate score on the second test was

80. This is in accord with the facts as regards difficulty of the two scales. Where the score for a single individual differed, the average was taken as the true score, but it will be noticed that this reduced the scores from the first scores on the average less than half a point. The scores were then changed into mental ages by means of the norms described above. That these norms would, in general, be preferable to age norms taken from the St. Paul Survey, say, is certainly clear to most persons who will read this article. One specific reason is given. The measures were made in St. Paul in the middle of the school year of pupils who had attained the age at which they are reported in most cases by the first of the preceeding September. They would thus have a better opportunity to overcome the hard steps between sentences three and four, and four and five on the Trabue scales than would the children measured, in the main, soon after the attainment of the age at which they are reported. The following is the correlation table by the Pearson formula:

	Mental	Ages	Deviations					
Cases	Yerkes-Bridges	Trabue	x	y	xy	x^2	y^2	Excess of Yerkes-Bridges over Trabue age
1	8.4	7.7	−2.0	−1.7	3.40	4.00	2.89	.7
2	9.8	8.3	− .6	−1.1	.66	.36	1.21	1.5
3	11.4	10.9	1.0	1.5	1.50	1.00	2.25	.5
4	12.1	10.0	1.7	.6	1.02	2.89	.36	2.1
5	9.5	9.3	− .9	− .1	.09	.81	.01	.2
6	9.8	8.3	− .6	−1.1	.66	.36	1.21	1.5
7	8.9	7.9	−1.7	−1.5	2.55	2.89	2.25	.8
8	11.2	7.6	.8	−1.8	−1.44	.64	3.24	3.6
9	12.8	13.0	2.4	3.6	8.64	5.76	12.96	− .2
10	9.7	7.3	− .7	−2.1	1.47	.49	4.41	2.4
11	10.5	13.0	.1	3.6	.36	.01	12.96	−2.5
Av.	10.4	9.4	r	.66	0=1	.31	0=1.99	.96 Av.

The Trabue scales are simple to give, and the key makes it possible to score the results in an objective manner. A careful person of some scientific training can give the tests and score the papers just as accurately as a trained psychologist. Dr. Trabue reports in his monograph, *Language-Completion Scales*, that Dr. Scott of Northwestern University finds a correlation of .64 by the method of ranks

with a series of tests designed to measure the qualities thought to be necessary for success as efficiency engineers. Other workers since have more fully demonstrated the reliability of the Trabue scales as a measure of intelligence closely related to general intelligence. The rather high coefficient of correlation (r .66) reported here may then be taken (reasoning the other way around) to indicate that the selection and classification of the pupils in the ungraded room is fairly efficient. The author believes that other cautious superintendents may wish to resolve some of the doubts they now entertain as to the basis in mental testing of the ungraded rooms they have established, or are establishing, in a somewhat similar manner.

In partial explanation of the generally higher level of the mental ages found by the Point Scale, it might be added that an examination of the detailed records of pupils on the several tests shows that every case, except case eleven, tends to do a little less well where the test requires most language ability, and best where most thing-thinking is required. This tendency is extremely well marked in cases 4, 8, and 10, and re-enforces the familiar recommendation of hand work, and more hand work, for the pupil in the ungraded class.

Sup't of Schools, WM. L. CONNOR.
Republic, Michigan.

REPORT OF NEW CASES AND MORE RELIABE AGE NORMS OF INTELLIGENCE BY THE POINT SCALE FOR THE BLIND

Since the publication of *Mental Measurements of the Blind,* (Psychological Monographs, Number 89, April, 1916), a number of blind pupils in the state schools of Ohio and Kentucky have been rated by the Point Scale for the Blind.

Instead of the scores of only seventy-eight persons either totally blind or with such defects of vision that none could see the small (one inch) black wooden cubes used in weight discrimination when placed before him on white paper on the table at which he was sitting, I now have scores of one hundred and sixty such persons. Furthermore, the numbers of VII, VIII, and IX year-old children are greatly increased, a point of serious defect in Table II of the Monograph.

The Table of Scores, Table I, page 147, of the *Journal of Educational Psychology* for March, 1916, included scores of each pupil in the original mental survey of the Ohio State School for the Blind whose vision was too poor to enable him to be fairly tested by the Yerkes-Bridges Point Scale. Their visual acuity was not equal to

Point Scale Scores. One Hundred and Sixty Blind Persons with Vision Too Poor to See Black Weight-Discrimination Cubes on White Paper on Table at Which Person Was Sitting. July 1, 1918

1	VI	VII	VIII	IX.	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII	XIX	XX	XXI	XXII	XXIII
	18	21	7	25	7	15	12	47	38	16	40	46	61	53	78	59	66	
		24	8	28	28	21	14	60	53	39	57	74	76	76	80	61	70	
		26	11	29	35	46	37	60	68	61	65	78	81	87	81	73	93	76
		28	12	34	36	47	44	64	74	62	69	83	87	90	84	83	97	81
		29	19	36	42	57	47	73	78	65	74	85	88	94	91	89	98	
		38	27	41	44	59	48	76	83	66	80	89	92		97	90		
		38	29	46	45	59	53	77	85	74	86	93	94		99	91		
		43	34	47	49	59	58	90	87	87	87		95			93		
		44	34	52	49	69	64		90	88	96							
		50	34	55	52	83	81		91	90								
			35	55	59	87	87		93	90								
			39		62		88											
			40		64													
			49		67													
			51		68													
			52		72													
			53		73													
2	18	33	34	41	49	59	50	68	83	66	74	83	87	87	84	86	93	
3	VI & VII 29		VIII & IX 34		X & XI 58		XII & XIII 60		XIV & XV 76		XVI & XVII 79		XVIII & XIX 87		XX & XXI 84			
4		34		46		57		76		71		85		85		89		
5		44	52	49	62	59	81	68.5	90	87	86.5	80.5	88	88.5	86	86.5	93	78.5

(1) Age nearest birthday when tested. (2) Medians 160 scores. (3) Medians by two year periods 160 scores. (4) Medians by two year periods 160 scores. (5) Medians by scores of 78 persons of Table II page 26 Monograph.

seeing the *pictures* well. But many of these persons could see the blocks quite well enough to perform the Knox Cube Test with no handicap. Such persons have different mental furniture from those who can merely see light and from the totally blind.

These new distributions of scores of the larger number of really blind pupils, which also reach down into the seventh year of life, with strength, give rise to a series of tentative norms which, from the numbers and distribution, are entitled to supplant all earlier presentations.

To the seventy-eight cases of Table II, page 26, *Psychological Monograph*, No. 89, the following additions have been made. Examiner, school, and date are given in each case.

Twenty-six	cases	T. H. Haines, Ohio State School for the Blind	Feb.	1916
Eleven	" (11)	T. H. Haines, Ohio State School for the Blind	Dec.	1916
Thirteen	" (13)	Miss A. C. Bowler, O. State Sch. for the Blind	Jan.	1917
Twenty-two	" (22)	T. H. Haines, Ken. Inst. for the Blind	Feb.	1917
Fourteen	" (14)	Miss A. C. Bowler, O. State Sch. for the Blind	Fall	1917

The same test material, examination sheets, and method of administration and scoring of tests have been used as described in the *Monograph*. Also the same method of distribution of scores by age.

The medians are given, as the most reliable tentative norms, especially for small numbers.

To show the improvement in conformity with the curve expected, if this point scale is a measure of intelligence, the medians, by years, of the seventy-eight cases of Table II of the Monograph are given in the last line of the table.

Both the series of medians by two-year periods show the expected advance during the years which intelligence tests from Binet onward have shown to be the years of marked development of that we call intelligence, which is measured by such tests.

The best norms in points, for the scale, suggested by this work to date, are as follows:

Years of Age	Expected Points
VI	18
VII	33
VIII	34
IX	41
X	49
XII	57
XIV	76

Ohio State University. THOMAS H. HAINES, M. D.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except July and August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated.

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of ten monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X MARCH, 1919 NUMBER 3

EDITORIAL

MENTAL TESTS AND COLLEGE ENTRANCE

The traditional college entrance examination has long been attacked as unsatisfactory by preparatory school teachers, by the candidates for entrance, and by college teachers. The theory of the examination was that a barrier was thereby set up to exclude from college those who were unprepared to do the work expected of them. Such expectation was capable of more precise definition when the bulk of college work consisted of Greek, Latin and mathematics, and when the preparatory school confined itself to the well established rudiments of these subjects. But with the introduction of electives into the college course and the later development of the group system, with the corresponding enrichment of the course of study in preparatory schools, and above all with the marvellous growth and increasing influence of the public high schools, it became more and more difficult to determine just what was to be expected of college freshmen and how their ability to meet these vague expectations could be revealed. The

preparatory teacher, especially the high school teacher, complained that the examination fettered him and prevented the kind of education which the child most needed. The high school graduates and their parents complained that the examination was arbitrary and capricious and did not allow the candidate to do himself justice. College teachers complained that under skillful coaching a very inferior student could be crammed sufficiently to pass the examination, and that classes were thus loaded with undesirable material. Thus any device that promised a broader and more reliable indication of the candidate's intellectual ability was likely to meet with a favorable reception.

Before the war tentative efforts had been made in a number of institutions to apply mental tests to freshman students, and to determine the correlation between the results of such tests and classroom work. The outcome of these efforts was not altogether satisfactory. Whether on account of the nature of the tests used or on account of the faulty technique in their administration, there was a distinct skepticism as to their reliability. The occurrences of the past two years, however, have given a tremendous impetus to applied psychology. The spectacular employment of mental tests on a large scale in the army has prepared the public mind for their use in other fields, and the necessity for selecting men for specific types of work has resulted in the development of tests that are said to have a high degree of reliability. For some time the Carnegie Institute of Technology has been using mental tests as an aid to the determination of the mental qualifications of freshmen, and now Columbia University announces that the traditional examinations will be supplemented by psychological tests to determine the candidate's ability to do college work. Professor Thorndike and his staff are at work on a battery of tests that will differentiate high school graduates into at least twenty grades or degrees of ability, and these tests will be available for other institutions after July 1.

All this is interesting and praiseworthy from the point of view of the scientific study of education, and it is to be hoped that the undertaking will receive the support that it deserves. At the same time the opinion may be hazarded that mere determinations of ability on the strength of general intelligence tests lack that precision and definiteness toward which scientific education is striving in diagnosis and prognosis. Such tests need to be supplemented by standard tests and scales in each of the high school subjects, so that the college authorities may have a complete diagnostic picture not only of the candidate's general intelligence (a vague term at best), but also of his proficiency in the work that he has attempted to do.

J. C. B.

NOTES AND NEWS

A correspondence course is announced by the University of California under the auspices of the California Children's Year Committee on the imagination, temperament, human cravings and play activities of children under twelve years of age. The material dealing with mental training has been prepared by Dr. Warner Brown, that on moral training by Professor C. E. Rugh.

Professor E. L. Thorndike, in his vice-presidential address before Section H, Anthropology and Psychology, American Association for the Advancement of Science, on "Scientific Personnel Work in the Army," recently published in Science, makes the following significant comments on the work of applied psychology: "Applied psychology is much more than cleverness and common sense using the facts and principles found in the standard texts. It is scientific work, research on problems of human nature complicated by conditions of the shop or school or army, restricted by time and labor cost, and directed by imperative needs. The secret of success in applied psychology or human engineering is to be rigorously scientific . . . . When we took pains to compute the reliability coefficients of all our data before going further with them, we saved time in the long run. Every failure to check apparent meanings by objective correlations was disastrous. An unverified hypothesis may possibly be a relatively harmless luxury if all one does with it is to think; to act on it is a grave danger . . . . . . It is relatively easy to be scientific when you can direct your talent in any one of ten thousand directions; yourself asking the questions for which you proceed to find answers! Psychology applied to the complicated problems of personnel work represents scientific research of the most subtle, involved, and laborious type."

Dr. Otis W. Caldwell, of the Lincoln School, New York City, and Mr. Stuart A. Courtis, of Detroit, have discovered a survey of the schools of Boston, Mass., made in 1845, and have printed an edition of 50,000 copies of questions in American history, geography, arithmetic and other subjects from the tests used in the survey. They desire to enlist the co-operation of the schoolmen of the country in giving these tests to eighth grade pupils some time during the last month of the present school year. The cost for a class of 30 pupils including record sheets and postage both ways will be $.65. The time required for giving the tests will be about 100 minutes of class time. Those who are willing to participate in this investigation are asked to write at once to S. A. Courtis, 51 Shelby Street, Detroit, Michigan. A certain number of the tests will be distributed free of charge.

At the February meeting of the New York Society for the Experimental Study of Education Mr. S. A. Courtis, director of the bureau of educational research, Detroit, Michigan, gave an illustrated lecture on "The Development of Standard Tests in Education, and Their Application to Problems of Classification and Diagnosis."

The new Prussian socialist Kultus-Minister has already introduced sweeping changes in the German school system. Among the 32 points of his program, as published by the *London Times*, there is to be complete separation of religion and religious authorities from the schools, all class schools are to be abolished and the "uniform" school maintained, co-education is more generally introduced, the teaching of history is re-organized, discussions of social policy are encouraged instead of forbidden, examinations are to be reduced, and teachers are to have representation in the government and in school administration.

The New York Branch of the American Psychological Association held a meeting on February 24, at which the following papers were read: "Correlations between the Standard Revision of the Binet Test and Performance Tests," by Dr. L. I. Stecher; "Examination for Emotional Instability," by Professor R. S. Woodworth; "The Mental State of Hysterics," by Professor H. L. Hollingworth.

Dr. Charles Barr Robertson, director of the division of university extension in the University of Pittsburgh, and one of the most popular school men in the State of Pennsylvania, died on February 4, after a month's illness.

Professor Alexander Inglis, of Harvard University, has been appointed by the Virginia State Board of Education to make an educational survey throughout the state.

Professor Irving King, of the University of Iowa, is now supervisor and organizer of the army convalescent school at Camp Travis, Texas. The Camp Travis Khaki College, a Y. M. C. A. school for soldiers, was organized during Dr. King's service in the southern cantonment.

Captain Garry C. Myers, Sanitary Corps, U. S. Army, is chief of educational service and morale officer at General Hospital No. 31, the former Indian School of Carlisle, Penna. Captain Myers is on leave of absence from the Brooklyn Training School for Teachers, New York City.

Captain R. H. Wheeler, professor of psychology in the University of Oregon, who has been conducting psychological tests in the army, has returned to take up his work at the university.

PUBLICATIONS RECEIVED

CARTER ALEXANDER. *School Statistics and Publicity.* Boston: Silver, Burdett and Company, 1919. Pp. xix, 332.

This book is intended for the school superintendent. It is designed to assist him in presenting the needs of his school system in the most effective and convincing manner. "The school superintendent must be a publicist. He must make reports to the public. In many places, for the next decade at least, he must fight as hard as any officer in the trenches to ward off the incessant and fierce attacks made upon his school appropriations by politicians and hardpressed but unthinking tax-payers. For warding off or beating back such attacks, his most effective weapons will be reports containing simple but skillful statistical devices for presenting the czaims of the school children." The book makes valuable suggestions for the collection of statistical data; indicates the significance of scales and tables of distribution; gives an admirably lucid discussion of central tendency, of deviation, and of correlation; and devotes almost one hundred pages to the discussion of devices for presenting statistics to the public in the most striking and attractive form. The wide-awake superintendent will find it extremely helpful and suggestive.

CHARLES R. ALLEN. *The Instructor. The Man and the Job.* Philadelphia: J. B. Lippincott Company, 1919. Pp. vii, 373. $1.50.

The instructor here referred to is the industrial instructor, whose business it is to take the crude, untrained, raw labor recruit in the factory and gradually develop such skill and efficiency as the man seems capable of attaining. This task has usually been left to the foreman, who has acted as though it were best accomplished by bullying and physical force. But managers and employers are coming to realize that this is a very wasteful method of procedure, and are turning to the specialist in handling men for more efficient methods. The author was formerly assistant superintendent of training in the United States Shipping Board, and has here set forth the results of his observations and experience.

WILLIAM H. ALLEN AND CLARE KLEISER. *Stories of Americans in the World War.* New York: Institute for Public Service, 1918. Pp. 176.

The authors have gathered together interesting personal accounts, newspaper reports, official citations, poems, and other material to give a composite picture of the part played by America in the war. It is excellent supplementary reading for the upper elementary grades.

WILLIAM H. ALLEN AND CARROLL G. PEARSE. *Self-Surveys by Teacher-Training Schools.* Yonkers-on-Hudson: World Book Company, 1917. Pp. xvi, 207. $2.25.

Critical self-analysis has not been a marked characteristic of American education. Especially the important function of teacher-training has been allowed to drift on the uncertain waves of tradition and local caprice. The present book is superficial and impressionistic, but it gives many valuable hints to live training school principals, and points the way to a more thorough and scientific study of the training of teachers.

J. Mace Andress. *Health Education in Rural Schools.* Boston: Houghton Mifflin Company, 1919. Pp. xii, 321. $1.60.

The poet and the back-to-the-farm enthusiast may laud the superior health-giving influences of country life as compared with city life, but every careful study that has been made shows the lamentable neglect of the simplest rules of hygienic living in our rural populations. The physical examinations for army service showed that the country boy was just as defective as the city boy, and country school children have little attention paid to their health. The present volume is designed to adapt accepted hygienic teachings and practices to the needs of the rural school teacher. It is written in a simple, direct style, yet is thoroughly scientific and up-to-date, and contains valuable lists of supplementary books. The final chapter presents an interesting point scale for the measurement of the teacher's work in hygiene, and the application of the scale to typical teachers one of whom gets a score of 16 points, another a score of 84 points, and the third a score of 133 points. The general use of this scale would give us some interesting data on health teaching in the schools.

Sholam Ash. *The God of Vengeance, A Drama in Three Acts.* Boston: The Stratford Company, 1918. Pp. xiv, 99. $1.00.

A translation of a Yiddish drama portraying the struggle of ideals in the heart of even the most debased. The scene is laid in a brothel in the Jewish quarter of a Russian town.

Frank P. Bachman and Ralph Bowman. *The Gary Public Schools. Costs.* New York: General Education Board, 1918. Pp. 86, with thirty infolded tables. Twenty-five cents.

One of the most thorough and elaborate studies of school costs that we have available. The Gary school shops have been widely claimed to be self-supporting. It was soon found that such a claim could rest only on a very confused and irregular system of accounting. The surveyors were obliged to go back to the original vouchers and to evaluate the output of the shops for several years. As a result they conclude that if all the shops be taken together they are only 54 percent. self-supporting, while if only the essentially productive shops are considered by themselves there are 69 percent. self-supporting.

Franklin Bobbit. *The Curriculum.* Boston: Houghton Mifflin Company, 1918. Pp. viii, 295. $1.50.

Social,commercial, industrial and political life are changing very rapidly, but our educational program lacks thorough-going adjustment to these changes. What should constitute the course of study for the different grades from the kindergarten to the college, and what should be the relative proportions of the subjects included? "For a long time we have been developing the theory of educational method, both general and special, and we have required both teachers and supervisors to be thoroughly cognizant of it. Recently, however, we have discerned that there is a theory of curriculum formation that is no less extensive and involved than that of method, and that it is just as much needed by teachers and supervisors." The present volume offers an introductory survey of the problems involved in curriculum making. There is a preliminary discussion of the ends to be served by the curriculum, a study of the technical training necessary for occupational efficiency, an analysis of education for citizenship, an examination of the bases of education for physical efficiency, an inquiry into education for leisure, and a consideration of the linguistic demands of social intercourse. One of the most interesting and stimulating chapters in the book is that on reading as a leisure occupation.

EMORY S. BOGARDUS. *Essentials of Social Psychology.* Los Angeles: University of Southern California Press, 1918. Pp. 159.

This is a kind of outline or syllabus of problems in social psychology as developed in the author's classes. There are chapters on the psychological basis of social studies, the instincts, the characteristics and operation of imitation, temporary and permanent groups and their influence on behavior, and the psychology of invention, leadership and social control. Each chapter is followed by an extended list of problem-questions and a selected bibliography.

SANGER BROWN, II. *Sex Worship and Symbolism of Primitive Races.* Boston: Richard G. Badger, Pp. 145. $3.00.

This simple presentation of a fundamental motive should appeal to everyone who is interested in mental evolution. The historical portion of the book gives a description of sex worship which had its origin in the mind of primitive man, but which continued its influence (unrecognized for the most part) through the past ages down to the present day. A parallel is drawn between the history of this motive in the collective mind of the race and the influence of the sex motive in the life of the normal individual.

ANTON CHEKHOV. *Nine Humorous Tales.* Boston: The Stratford Company, 1918. Pp. ix, 60. Twenty-five cents.

These delightful tales are a relief from the sinister, dark, morose character of most modern Russian fiction. They remind one strongly of O. Henry in the directness of their style and the whimsical nature of their development.

HUBERT GUY CHILDS. *An Investigation of Certain Phases of the Reorganization Movement in the Grammar Grades of Indiana Public Schools.* Bloomington, Ind.: Indiana University Book Store, 1918. Pp. viii, 187. $1.50.

This is an investigation of the Junior high school, based upon a study of thirty-five so-called junior high schools, thirty-five typical departmental schools, and twenty-three non-departmental schools. Comparisons were made as to standards, programs of study, provision for individual differences, revised methods, social organization, costs, achievements in specific subjects, and retention of pupils in school. In many respects the junior high schools show improvement over the elementary schools in flexibility of promotion, work in practical arts, and wider opportunities for electives. The Woody Arithmetic Tests, Series B, Multiplication, the Ayres Spelling Scale, the Thorndike Reading Scale, Alpha 2, and the Thorndike Visual Vocabulary Scale were given to the pupils of schools in two counties which differed widely in the time given to English and arithmetic. Except for the spelling, the schools giving little time to these subjects made practically as good a showing as those that gave twice as much time to them.

SHERWIN CODY. *Commercial Tests and How to Use Them.* Yonkers-on-Hudson: World Book Company, 1919. Pp. vii, 216. Ninety-nine cents.

This little volume gives an account of the origin and use of the National Business Ability Tests. The chapter on "Principles of Scientific Tests" has a rather full account of the Alpha Army Tests and the appendix contains a comparison of the performance of the same pupils with arithmetic tests of the National series and the Courtis Tests, Series B. There are detailed directions for giving and scoring the tests, and the results of their application in various communities.

Maxim Gorki. *Stories of the Steppe.* Boston: The Stanford Company, 1918. Pp. vii, 59. Twenty-five cents.

This is a volume of "The Universal Library" in which the publishers present the best literature of all nations in neatly bound volumes at twenty-five cents each. The series already contains eight or ten numbers. The present volume contains three of Gorki's vivid and poignant short stories.

Samuel Hamilton. *Hamilton's Standard Arithmetics.* Cincinnati: The American Book Company, 1918. Book I, Pp. 256+xvi; Book II, Pp. 300+xx: Book III, Pp. 374+xxii.

These texts are generously illustrated, are well supplied with devices for drill and short cuts, and offer an abundance of practical problems.

Alice M. Krackowizer. *Projects in the Primary Grades.* Philadelphia: J. B. Lippincott Company, 1919. Pp. ix, 221. $1.28.

The traditional elementary school is admittedly formal and artificial. Yet abundant materials for real education lie all about children waiting the proper organization. In this book the author makes happy suggestions for utilizing the child's play activities, his social experiences, and his interest in nature, in building up preliminary habits for the more formal subjects of the school. Teachers of small children will find the book very helpful.

Paul J. Kruse. *The Overlapping of Attainments in Certain Sixth, Seventh and Eighth Grades.* New York: Teachers College, Contributions to Education, No. 92, 1918. Pp. 91.

This is an intensive study of the behavior of a limited number of pupils to a large number of mental and educational tests. The 860 pupils were distributed fairly evenly among the three grades, and the tests used were the Woody Arithmetic Tests, three arithmetical problem tests, four Trabue Completion tests, two composition tests, Thorndike Reading Scale Alpha 2, a spelling test, four opposites tests, seven other association tests, three directions tests, and two visual vocabulary tests. The author finds great overlapping from the results of single tests, but this overlapping is reduced materially in the composites of a number of tests of the same trait, and still further reduced in composites combining the measures of different traits. The coefficients of reliability of the arithmetic and English composites were very high, approximately .85.

Arthur W. Kallom. *Arithmetic. Determining the Achievement of Pupils in Common Fractions.* Boston: Department of Educational Investigation and Measurement, Bulletin No. 15, 1918. Pp. 38.

This bulletin reports the results of investigations extending the previous study of addition of fractions to multiplication and division. An analysis was made of the types of mental processes involved in multiplication and division of fractions, and tests were constructed to reveal proficiency in these processes. The tests were given to 3513 pupils of grades VI–VIII in 95 classes. The results are tabulated to show the medians for speed and accuracy in each test for each grade, the percentage of failures in each test, and the sources of error in the attitude of the pupils. The bulletin closes with a valuable plan of diagnosis to be used by the teacher

Richard Swann Lull, Editor. *The Evolution of the Earth and its Inhabitants.* New Haven: Yale University Press, 1918. Pp. xi, 208. $2.50.

This is a series of lectures delivered before the Yale Chapter of the Sigma Xi by specialists in various fields of evolution. It includes "The Origin of the Earth"

by Joseph Barrell; "The Earth's Changing Surface and Climate," by Charles Schuchert; "The Origin of Life," by Lorande Loss Woodruff; "The Pulse of Life," by Richard Swann Lull; and "Climate and the Evolution of Civilization," by Ellsworth Huntington. The book presents an excellent survey of the most recent findings of various sciences on the subject of evolution.

INEZ N. McFEE. *The Teacher, the School and the Community.* New York: The American Book Company, 1918. Pp. 256. $1.24.

This book is intended as a "guide, philosopher and friend" for the new teacher in the ungraded rural school. For the graduate of one of our better normal schools or training schools it would be entirely superfluous, but unfortunately there are thousands of young teachers all over this country who have had so little preparatory training that the book would be a godsend to them. There are numerous references to teachers' helps that will prove valuable to those seeking such material.

ALBERT E. McKINLEY, CHARLES A. COULOMB, AND ARMAND J. GERSON. *A School History of the Great War.* Cincinnati: The American Book Company, 1918. Pp. 192.

Already we are far enough away from the great tragedy to make a school history of it very welcome. The present study is an admirable story of the war and will make a strong appeal to young people. The narrative is brought down to the signing of the armistice on November 11, 1918.

BRANDER MATTHEWS. *An Introduction to the Study of American Literature.* Cincinnati: The American Book Company, 1918. Pp. 268.

A new and revised edition of this valuable little manual. The author still holds to the principle of excluding all living writers, thereby securing more space for the consideration of the earlier literature, but robbing his readers of the benefit of his opinions on contemporary movements.

EDWARD J. MENGE. *The Beginnings of Science, Biologically and Psychologically Considered.* Boston: R. G. Badger, 1918. Pp. 256. $2.00.

This series of papers presents a Roman Catholic reaction to the trends of modern science. There are treatises on biology, psychology, genetics, metaphysics, logic, evolution, vitalism and the ideal. The author's interpretations are scarcely in accord with the views of most leaders in scientific thinking.

WALTER SCOTT MONROE. *Measuring the Results of Teaching.* Boston: Houghton Mifflin Company, 1918. Pp. xviii, 297. $1.60.

This is perhaps the best presentation of the results of educational measurements to date. The book embodies much of the material in the author's previous *Educational Tests and Measurements,* written in collaboration with Mr. James C. DeVoss and Dr. Frederick J. Kelley, but omits many of the tests mentioned in that book, and adds some others. The omissions enable the author to give a more extended treatment to the problems of diagnosis and to the correction of the defects revealed by the tests. On the other hand the emphasis on a few types of tests leaves the book unsatisfactory as a comprehensive account of the present status of educational measurements. In the very beginning, however, the author disavows any intention of producing a fundamental treatise on the subject, but has aimed rather at "a text which will help teachers to use standardized tests to the greatest advantage." In the chapter on reading only the Monroe Silent Reading tests, the Courtis Silent

Reading test, the Thorndike Visual Vocabulary, and the Gray Oral Reading test are discussed, but over 50 pages are devoted to the correction of defects in reading. The appendix contains valuable suggestions for procuring the most important standard tests.

CHARLES CLINTON PETERS. *Human Conduct. New York:* The Macmillan Company, 1918. Pp. xii, 430. $1.30.

The sub-title of this book is *A Textbook in General Philosophy and Applied Psychology for Students in High Schools, Academies, Junior Colleges, and for the General Reader.* Many educators have been asking, "Why should high school science be limited to physical science? Why would not mental science offer valuable educative material to the adolescent?" There is no essential reason, other than tradition, for the restriction of psychology and sociology to college students. Already progressive schools are introducing courses in economics. The present book selects those phases of psychology and logic which would make the strongest appeal to young people and sets them forth in direct and simple language. The first part of the book is concerned chiefly with illusions, and applies the psychology of illusions to an explanation of the way in which misunderstandings arise. Among other topics discussed are apperception and its significance for the individual and the race; the solution of problems and clear thinking; the search for the causes of things; the control of conduct through ideas; the role of memory and imagination; attention and habit formation; and the selection of one's life work. While the book is somewhat diffuse and discursive, the reading of it would be of distinct benefit to any high school pupil.

JOHN C. STONE. *The Teaching of Arithmetic.* Chicago: Benjamin H. Sanborn and Company, 1918. Pp. v, 262. $1.32.

This is a very clear and simple presentation of the essential facts of arithmetic. The first chapters deal with the fundamental number combinations, chapter eight suggests many games that may be utilized in drill work, and subsequent chapters discuss common and decimal fractions, percentage, denominate numbers, analysis and solution of problems, and planning the work in arithmetic. The final chapter is entitled "Measuring Results," and contains a summary of the work of Rice, Stone, Courtis, Woody, and the tests used in the Cleveland Survey.

C. W. TABER. *The Business of the Household.* Philadelphia: The J. B. Lippincott Company, 1918. Pp. xii, 438. $2.00.

This volume is in keeping with the high standard of excellence set by previous volumes in the Lippincott's Manuals. Part I deals with fundamental principles of household finance, including the family income, bank and household accounts and the family budget. Part II discusses the necessities of the budget, such as rent, fuel, light, taxes, insurance, and marketing. Part III considers such problems as service in the home, insurance, savings, investments, old age, and the cultural wants of the family. Part IV takes account of the legal and business status of the family, treating of real estate titles, inheritances and wills, laws affecting the family, and business principles in the home. The discussions are thorough and scholarly, and based upon the best authorities.

ORDWAY TEAD. *Instincts in Industry.* Boston: Houghton Mifflin Company, 1918. Pp. xvii, 222. $1.40.

This is not a psychological study, but an attempt to analyze and present certain factors of the labor situation on the basis of a scheme of instincts. The author

selects what he considers ten basic instincts, and under each rubric calls attention to certain crucial problems in industrial life. What he has to say about industrial problems is interesting and pertinent, but for the psychologist who is at all critical in the use of his terms the attempt to explain anything by the employment of such terms as the instinct of workmanship, the instinct of self-assertion, the instinct of submissiveness, and the instinct of the herd is pure word juggling. The book is another illustration of the pernicious effect upon economic writers of the craze for instincts started by James and fostered by McDougall and other social psychologists.

FREDERICK J. TEGGERT. *The Processes of History.* New Haven: Yale University Press, 1918. Pp. ix, 162. $1.25.

This is an attempt to do for human history what biologists are engaged in doing for the history of the forms of life. The great problem of history is to determine how man everywhere has come to be as he is. The various general theories that have been proposed are carefully examined, the geographical factor in history is given due consideration, and the age-long conflict between individualism and collectivism is shown to be the fundamental human element in history. The book is particularly valuable for its point of view and for the method of attack proposed for the problems of history.

WILLIS MASON WEST. *History of the American People.* New York: Allyn and Bacon, 1918. Pp. xxvi, 729+44. $1.75.

This text, intended for use in high schools, emphasizes the development of industrial and social life in America. According to the author the following points have received special attention: (1) The historical grounds for friendship between America and England, in spite of old sins and misunderstandings. (2) The meaning of the West in American history. (3) The heroic labor movement of 1825–1840, usually ignored. (4) The long conflict between intrenched "privilege" and the "progressive" forces in State and Nation. A final chapter on the World War brings the narrative down to January, 1918. The book is profusely illustrated with pictures, maps and diagrams, and the liberal use of paragraph numbers and headings and bold face type brings out the statements which the author considers of prime importance.

JOHN F. WOODHULL. *The Teaching of Science.* New York: The Macmillan Company, 1918. Pp. xiv, 249. $1.25.

This collection of papers and addresses touches upon a large number of topics connected with the teaching of science, and covers a period of over twenty years. Among the subjects discussed are the enrichment of high school courses in physics and chemistry to give them greater cultural content and significance, the evils of specialization in high school science, the effects of college entrance requirements in natural science, learning from experience, and the project method in the teaching of science. In some measure the book serves as a history of the natural science movement in recent education.

VOLUME X APRIL, 1919 NUMBER 4

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE RESPONSE OF A COMPOSITE GROUP TO THE STANFORD REVISION OF THE BINET-SIMON TESTS

H. W. CHASE AND C. C. CARPENTER
University of North Carolina

During the winter and spring of 1916–1917, the writers tested 103 children in the elementary school of Chapel Hill, N. C., by the Stanford Revision of the Binet Tests. After some preliminary training work, the actual testing was done by one of us (C. C. C.). Terman's directions, as laid down in his *Measurement of Intelligence,* were followed absolutely, and conditions were favorable in all ways throughout.

For the purposes of this study, the records of all children below 9 years 0 months and above 12 years 11 months were disregarded. This leaves a total of 77 records, comprising those of all children in the school whose chronological ages ranged from 9 to 12 years inclusive. Most of these children were in the fourth, fifth, and sixth grades.

The relevant facts as to the distribution of the intelligence quotients of this group of 77 children are shown in Table I.

It will be seen that our distribution is somewhat skewed. Our results differ throughout from Terman's in that the median mental and median chronological ages fail to coincide at any of our age-levels. For children of the ages tested by us, Terman reports median I. Q.'s as follows: 9 years, 100.5; 10 years, 103; 11 years, 98; 12 years, 98.[1] The average of his medians is, for the four years under consideration, 99.9. Our medians for these ages have been stated in Table I; their average is 92.2 and only 23% of our children reach or exceed the average of his four medians. The average of the medians of our two larger age-groups (10 and 12 years) is 91.9.

The divergence of our results from Terman's is too considerable

[1]*The Stanford Revision and Extension of the Binet-Simon Scale for Measuring Intelligence.*

to be dismissed as due to possible variations in our technique from that prescribed by Terman. It is to be explained rather by the

TABLE I

DISTRIBUTION OF I. Q. OF 77 CHILDREN FROM 9 TO 12 YEARS

	Age 9	Age 10	Age 11	Age 12	All
No. tested	11	26	12	28	77
Range of I. Q.	90.4–118.4	68.8–140.7	69.6–119.2	58–123.8	58–140.7
Middle 50%	94.3– 99.9	88.1–103.2	75.3– 98.2	83.2–97.3	85.2–99.7
Median	98.2	93.1	86.7	90.7	92.2
No. with I. Q. below 70	0	1	1	1	3
No. between 70-80	0	0	3	5	8
No. between 80-90	0	10	3	7	20
No. between 90-100	7	7	2	12	28
No. between 100-110	3	5	1	0	9
No. between 110-120	1	1	2	2	6
No. between 120-140	0	1	0	1	2
No. above 140	0	1	0	0	1

composition of the group itself. Of the 77 children, 12 are members of families connected professionally with the University of North Carolina. The intelligence quotients of this group of children are 100, 103.2, 106.3, 106.5, 110.6, 112.1, 118.4, 119.0, 119.2, 120.3, 123.8, 140.7. The median is 115.25. All the children, it will be noted, are above the average of Terman's medians for the indicated years. In terms of his interpretation of the significance of intelligence quotients,[2] four of these children are of average intelligence (90–110), five of superior intelligence (110–120) and three of very superior intelligence (120–140); one of the three, indeed (140.7) reaching a level attained by only one child in 250 or 300.

The remaining 65 children form a rather homogeneous group, descended in the great majority of cases from stock which has lived for generations in this or in closely similar localities. Many of the families are inter-related, and all have in general the same standards, traditions, and a common range of experiences and ideas. The fathers are small business men, skilled workmen, etc., and small farmers living in the town or in the adjacent country districts. Thirteen of this group live outside the town limits, and had attended

[2] *The Measurement of Intelligence*, p. 79.

the school at the time when the tests were made for a median time of two years. They are thus somewhat less favored educationally than the town group, as the country schools from which they came are without exception poor. The intelligence quotients of this "country" group of 13 are 69.6, 73.1, 75.3, 75.6, 76, 77.7, 81.5, 88.1, 88.2, 88.3, 89.2, 94.2, 96.6, with the median at 81.5. No one of these children, it will be noted, reaches 100. The faculty" and the "country" groups are thus completely separated by the tests, the median of the latter group being 33.5 below that of the former.[3]

Of the remaining 52 (the "town" children), the median I. Q. is 92.3. Two have I. Q.'s of less than 70 (58, 68). Four exceed 100, the highest record being 111.4 (mother a teacher).

On the basis of the data given by Terman in Appendix 1 of the "Stanford Revision," the attempt was made to discover whether the deficiency of the 65 town and country children was general, or was localized in certain tests. The data obtained by Termna, Lyman, Ordahl, etc., by Terman and Childs, and by Terman, Trots and Waddle, were utilized in the attempt to discover how the percentage of children of a given age passing a given test compared in Terman's investigations and in our study. As our age-groups are small, and as the percentages of children of a given age passing a given test vary, sometimes considerably, in the different investigations reported by Terman, any difference between Terman's figures and our own for children of any one age in any one test naturally means little or nothing. It may be assumed, however, that any marked deficiency in our children of all four ages in any one test will possess some significance. Such deficiencies exist in tests 10–1, 10–3, 10–6, 12–1, and 12–2. In other words, children of all four ages tested by us were deficient in vocabulary (10–1, 12–1), in drawing designs from memory (10–3), in the giving of sixty words in three minutes (10–6), and in the ability to define abstract words (12–2). The deficiency in 10–3 (drawing designs) is possibly due to the fact that instruction in drawing is not given in the school. The other deficiencies suggest at least that the children's stock of speech-habits is small, and that those which have been acquired are by no means in a condition of readiness to function. The results of practically all tests, however, were poor. In only two cases (making change and reading and report), were the scores of all our age-groups

[3]Eight other country children, not within the designated ages, were tested in the course of the work. Their I. Q. 's are 48.3, 53.4, 64.3, 67, 71.8, 75, 82, 83.7. No other faculty children were tested.

up to Terman's Standards. In view of this and of the facts noted above, it is doubtful whether much importance should be attached to the particularly bad records in the tests above mentioned. The records of the faculty children in these tests are good. In naming 60 words all passed except one child of three at nine years and one of four at ten years. At nine and ten, all passed the 10-year vocabulary, one at nine and two at ten passing the 12-year vocabulary also. At eleven and twelve, all passed the vocabulary tests for 12 and 14 years.

It is interesting to note that the reaction of our town and country group of 65 to the tests is considerably more homogeneous than that of the children of the same ages tested by Terman. The average deviation of the scores of the 362 children from 9 to 12 tested by him (see pp. 35 and 36 of the *Stanford Revision*) is approximately 10, ours is 7.5. This greater homogeneity is due largely to the fact that the I. Q.'s of 43.1% of the 65 children were bunched between 86–95, while only 33.9% of Terman's children are included within any 10-point range (that from 96–105). The percentage distribution of the I. Q.'s of these 65 children, as compared with Terman's distribution, is shown in Table 2. The greater homogeneity of our group is natural in view both of the interrelated stock from which the children come and of the similarity of their environment.

TABLE II

PERCENTAGE DISTRIBUTION OF I. Q.'S FOR 65 CHILDREN (CHAPEL HILL) COMPARED WITH DISTRIBUTION FOUND BY TERMAN

I. Q.	Terman (%)	Chapel Hill (%)
56–65	0.33	1.5
66–75	2.3	7.7
76–85	8.6	20.0
86–95	20.1	43.1
96–105	33.9	23.1
106–115	23.1	4.6
116–125	9.0	0.0
126–135	2.3	0.0
136–145	0.55	0 0

It is clear that we have in our group of 77 children three different levels of response to the test situations. The high-level response of the faculty group is in agreement with the findings of other in-

·The Terman percentages aref rom the *Measurement of Intelligence*, p. 66. They are his combined results for children from 5-14. These figures differ so slightly from those for children for 9-12 years that it is a matter of indifference which data are used for the comparison.

vestigators who have tested children of superior social status. The lower level of response of the town and country group is more difficult to explain. It is certainly not due to the fact that these groups are as a whole, or to any considerable degree, made up of children of "inferior" social status in the sense in which Terman and others have used the term. Judged by the standards of small-town communities throughout the state, most of the town children are clearly of average social status, and the same is true of the country children if judged by the prevailing standards of rural life. It is certainly not the case that the thirteen country children between the ages of 9 and 12, and the eight others without those ages, come from the inferior homes of their neighborhoods. It is not the children of the most ignorant and unprogressive farmers who are sent into town to school. And yet, of these 21 children, not one has an intelligence quotient reaching 100, and eight have intelligence quotients of 75 or less. Of the 52 town children, very few indeed are from homes sufficiently below the average of the typical small-town community to be called inferior, while a few are from homes distinctly above the average for their group. Yet the median intelligence of the group is 7 points lower than that of Terman's average group. (*Stanford Revision*, p. 91)

It may seem that to emphasize the fact that these children come from "average" homes is simply to quibble over words; that the only fact worth considering is that the home environment of those children, though average as measured in terms of their own fellows, may be inferior if judged by the standards of environment in which Terman's work was done. The case is, however, by no means so simple. From the average homes of any community, whatever measured in absolute terms their standards may be, come the judgments and opinions which largely regulate the community life. The inferior home, however great its social importance may be in other ways, is by no means equally important in shaping community standards and activities. Thus the social importance of a low or high level of intelligence in the average home is obviously very great. We are far from intimating that the intelligence quotients of children from average homes in the small towns and rural neighborhoods of this State, or of this section of the State, average lower than those of the Western children examined by Terman. We have shown only that such a condition does exist in children now in the upper grades of the elementary school in this community, excluding the "faculty" children. From the results of scattering tests made

in other grades, and from our observation generally, it seems probable that the condition exists throughout the school.

Our work has, however, gone far enough to convince us that, if the Stanford or any other intelligence scale is to be made available for use throughout the country, a great amount of work needs to be done in the derivation of norms. It seems to us probable that no general norms for children of inferior, average, and superior social status can be set up, but that sectional norms, perhaps norms for states or for homogeneous groups of states, will be needed. Furthermore, most measurements of intelligence have been made with city children. Can rural children be judged by the same norms? Or can they in some sections of the country and not in others? Is it possible to judge satisfactorily the intelligence of an individual without knowing first what normal intelligence means in the community from which he comes? For so long as he remains in that community, his successes and failures will be determined, not by any abstract standard of "average" intelligence, but by the relation of his mental ability to that of his neighbors.

Such questions become of great importance—to take but one illustration—in any attempt to use intelligence scales to diagnose feeble-mindedness. Terman, though recognizing that the boundary between normality and subnormality is "absolutely arbitrary," states definitely that "all who test below 70 I. Q. by the Stanford revision of the Binet-Simon scale should be considered feeble-minded, and it is an open question whether it would not be justifiable to consider 75 I. Q. as the lower limit of 'normal' intelligence."[5] Now of the 21 country children tested in all by us, 5 test below 70 I. Q. and 7 below 75 I. Q. Are these children all to be classed as feeble-minded? Two of them are clearly such. We seriously doubt whether the others can be so classified. They are not, to be sure, capable of progressing far in school; but, on the other hand, there is every reason to believe that, in the simple environment in which they will pass their lives, they will sustain themselves satisfactorily. When it is remembered that the median intelligence quotient of the country group is approximately 80, it is clear that individuals who test between 60 and 70 will be able to compete on terms which are not too unequal.

It is possible that we shall find that we need to distinguish between absolute and relative feeble-mindedness. A judgment of absolute feeble-mindedness would mean, say, that the individual

---

[5]*The Measurement of Intelligence*, p 81.

tested below some point set up on the basis of nation-wide or even inter-national experimentation with intelligence scales. It would signify a lack of that degree of intelligence necessary to enable the individual to adjust himself satisfactorily to the demands of life in typically "modern" communities, with their complexity of organization. Such absolute standards are of course the ones which need to be considered in such occupations as army service, in which the demands made on the intelligence of men are the same throughout the country. A judgment of relative feeble-mindedness, on the other hand, would mean that the individual was incapable of adjusting himself to the demands of life in his own, or in similar environment. The point below which it would lie would vary with many factors. It is, of course, possible to carry the argument to an absurd extreme, and to contend that such individuals as the Jukes should be considered "relatively" normal, since they get on well enough at their low level of existence. But such individuals are considered, even by their neighbors in the same general environment, as abnormal. The point is just that such individuals cannot adjust themselves satisfactorily to the life of even the simplest sort of decent and self-respecting modern civilized community. It is altogether probable, on the other hand, that individuals with intelligence quotients of between 60 and 70 *can* adjust themselves fairly satisfactorily to the demands which life makes on the average small farmer in the vicinity in which our tests were given. To the more complex life of a larger community such individuals would probably fail to adjust themselves; in this sense their normality or feeble-mindedness is relative.

Intelligence tests given to Southern negroes, again, will unquestionably show that, judged by an absolute standard, enormously high percentages are feeble-minded. But certainly one would miss the point altogether who should fail to remember that, in terms of their own environment and race, most of these individuals are normal enough.

But to return to the tests. The median I. Q. of the faculty group is, it will be remembered, 115, that of the town group is 92.3. This difference of 23.7 points is considerably greater than Terman found between his inferior and superior groups (14 points).[6] For the ages tested, it amounts to a difference of from two to three years in mental age. It is a difference as great as that found by any investigators between groups of children of unlike social status.

[6]*The Stanford Revision*, p. 91.

Thus, Yerkes, who tested children of an inferior and of a superior group in the Cambridge schools, estimates "the superiority of the favored individual over the unfavored as 20 per cent." by the point scale, and adds that this difference is doubtless too low for middle childhood and too high for early adolescence.[7] In terms of mental age, this means a difference, at the ages tested by us, of from two to three years. With this result the work of Bridges and Coler in Columbus closely agrees.[8] It seems fairly certain, then, that American children of the ages tested by us may differ, because of unlike social status, by from two to three years in mental advancement.

Now does unlikeness in social status produce differences in the level of response to intelligence tests predominately because of unlike native endowment in the children tested or because of unlike environments in which they have lived? Yerkes, and Bridges and Coler, prefer not to raise the question, while Terman concludes that native endowment is the predominant factor.

It is clear that the favored groups, both in Cambridge and in Columbus, were from environments different in all important respects from those of the unfavored groups. Not only did the two sorts of children in each city come from homes of different type, but they had attended different schools, their free time had been spent with different companions, the neighborhood influences as a whole were different. The unfavored children tested by Yerkes, though all from English-speaking homes, lived in a quarter largely populated by foreigners. Bridges and Coler mention specifically the poor school building, inferior teachers, and bad neighborhood influences generally of their unfavored group.

Now faculty and town children tested by us differ far less from each other in their general environment than do the superior and inferior Cambridge and Columbus groups. Of the faculty children, seven were born in the town and all the remainder save one—who came from a closely similar environment—have lived in the town since before their school-days began. The two groups have attended the same school, learned from the same teachers, played together freely in and out of school, been cared for in their early years by the same type of colored servants, enjoyed the same pastimes, reacted to the same rural environment. In home life and in occasional travel only has the faculty group had a marked advantage.

[7]Yerkes, R. M. *A Point Scale for Measuring Ability,* Chap. 6.

[8]Bridges, James W., and Coler, Lillian F. *The Relation of Intelligence to Social Status.* Psy. Rev. 1917, *24,* 1-31.

In many important features the environment of the two groups has been practically the same. Yet they differ in their test records as greatly as do children from good and poor quarters of cities, who have little in common in the way of environment. That such differences should still exist when the action of unlike environmental influences is reduced as nearly to a minimum as it could well be in children of different social status suggests strongly that the records of individuals in tests like those making up the intelligence scales now in use are determined predominately by their native endowment. It is difficult to believe that the homes from which the faculty group comes are so superior in cultural influences to those of favored groups tested in other places as to compensate for the absence of other environmental differences.

We would not, of course assert that environmental influences have nothing to do with the results. On the contrary, the poor record of the unfavored children in such language tests as those of vocabulary and sixty words is clearly in part due to the fact that little reading goes on in the homes from which many of them come, and that outside of school they come less into contact with the world of ideas. But it should not be forgotten that the reaction of the unfavored children to the tests as a whole was poor; it would be difficult, for example, to explain on the ground of environment the fact that children of two of the four ages tested (10 and 11) were notably deficient in the ball and field test—surely if environment ever favors such a test, it does so in a small-town community.

If heredity is the predominant factor in the individual's reaction to the tests, the unusually good record of the faculty group is only to be expected, since the families from which they come owe their presence in the town to the fact that at least the father has been chosen as a member of the community because of excellence in mental qualities presumably closely similar to those tested by the intelligence-scale.

It might seem that the hypothesis that heredity is the main determining factor in the responses breaks down when one attempts to explain by it the difference between the town and country groups. These two groups are of homogeneous stock, and much intermarriage has taken place between their families. Differences in environmental influences do exist; the standards of living are somewhat lower in the country group, they have attended the town school for shorter periods and suffered from inferior teachers in their early school days, their environment as a whole has been less stimulating,

their attendance at school has been more irregular, and conditions as to health and sanitation in their neighborhoods have not been so good.

But, to one who is familiar with the locality, the influence of an heredity factor is clearly apparent. The farming land about the village is generally poor, and the rewards of the average small farmer's labor have been as a rule meager. The incentives for intelligent and ambitious individuals to remain in the locality have thus been slight. That there has been a drift of population away from the county, which contains only two towns, both of less than 2000 inhabitants, is evident from the fact that the population of the county, which was 14,948 in 1890, was only 15,064 in 1910. This drift has been both to the towns and to cities without the county or the state. Thus, there was, in the 20 years just mentioned, an increase in the village itself of 132. As the whole township of which the village forms part only increased in the same period by 80, it is clear that there was in the rural sections of the township an actual decrease in population during the 20 years. Some of this decrease is accounted for by migrations of more ambitious individuals to the village in search of better opportunities, part of it by migration to other localities for the same reason. The country neighborhoods have, in other words, been subject to an adverse selection, which has undoubtedly removed many of the more capable individuals. The poor environmental influences of the children would then, in large measure, have to be considered as symptoms rather than as causes. Though the town has profited to some extent by immigration from the country districts, it has, because of its rather restricted opportunities, been subject to the same adverse selection. Its small net increase in population during the 20 years up to 1910 has been mentioned. The fact that the faculty families have been attracted *to* the town because of intelligence above the average, while the more able "native" inhabitants of the town have been attracted *away* from it for the same reason, has then possibly produced the situation revealed by the tests. In this fact, rather than environmental differences, lies the explanation of the unlike responses of the groups.

It remains only to suggest that many thousands of small town and rural communities throughout the country have been subjected to this process of adverse selection. What has been the effect of this selection upon the capacity of the younger generation in such functions as those tested by the scale. The answer to this question would be of the very greatest social and educational importance.

AN EIGHTH GRADE DEMONSTRATION CLASS AND THE THREE R'S

COLIN A. SCOTT
Mount Holyoke College

The village school of South Hadley, Massachusetts, was brought, in 1917, into connection with the Education Department of Mount Holyoke College by a small grant from the college which was used in paying part of the 8th grade teacher's salary. The purpose was to make a school which would be able to demonstrate some of the principles of modern education and afford a means of observation and study for the students of the Education Department. It was thought best for the first year to concentrate effort on one grade and to demonstrate there what could be done by practical scientific analysis followed up by such methods as could be based on this analysis.

There was no attempt to suggest a system of preconceived methods for the teacher. Even the most philosophically founded principles of education are liable to prove abortive when handed over to the teacher in the form of methods. In the demonstration class, on the contrary, we were accustomed to say, "When success flatters or when difficulties arise, analyze the situation. Look for your methods afterwards." If this itself may be called a method, it is one of attitude. It does not consist of rules previously learned.

Arithmetic

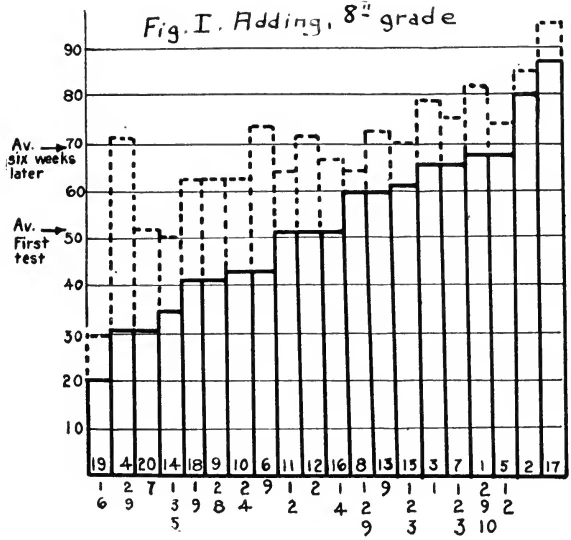
The situation in this 8th grade demonstration class which first attracted us was that of arithmetic. For administrative reasons, not necessary to detail, a beginning was not made until late in November. At this time the children were doing seat-work on Courtis blanks for adding, subtracting, multiplying and dividing, and were having some oral drill in adding, etc., from tables on the blackboard. This was done in adding, for example, by having one child at a time add while the other children were supposed to follow. It is obvious that if the child chosen is a rapid adder, the other children are unable to follow, while if he is slow, the others can hardly be interested enough to keep pace with him, or if they do the practice is useless to them. One person besides the teacher does the work. If the class is twenty in number, therefore, it ought to be

possible to find a way to have twenty times as much work done in the time. A situation very like this is found in many other subjects beside arithmetic where the traditional class teaching is the ideal pursued. As for the seat work on the Courtis blanks, it was found that although all the children were working, the practice did not seem to improve their speed to any extent. The difficulty standing in the way of the progress of each child had not been discovered. Under these conditions, mere repetition may result in perpetuating a habit essentially slow.

Besides the drill work in the four simple operations, work in problems was also going on. This was of the conventional character drawn from the book and consisted in offering the children problems to solve on the analogy of problems previously explained by the teacher. There was no opportunity provided for the children to frame problems for themselves on the basis of their observation and experience of things with which they came in contact. There was no stimulus therefore for them to inquire along quantitative lines. They did not ask questions. They merely answered them, and this naturally according to stereotyped models previously provided. Procedure of this kind is traditional and widespread, but since the inventive and inquiring attitude is absent, there is very little about it which can be said to exercise the reasoning powers.

The more careful and scientific analysis of the situation in arithmetic was begun upon the four simple operations, the other work going on meanwhile, as it had always been conducted. To economize space we will follow the analysis of adding, and omit the full description of the work done on the other three simple operations; but the reader will understand that similar work was done in these subjects also.

One reason for choosing the four simple operations to start with was that in the case of adding, for example, the work to be done to make oneself an efficient adder is simple and limited. It is simple because the basis of the operation requires only the immediate association of a pair of digits, given as the stimulus, to the digits representing their sum, with a background consciousness of the set, attitude, or task to be accomplished. The last factor is not one of great educational importance or difficulty, but it is nevertheless true that children in adding a list of digit pairs will sometimes give the multiplied product instead of the sum of the numbers before them. The task itself is simple. To see, for example, seven and nine, to read



these digits thus, pronouncing them mentally, and immediately to respond by saying, either aloud or mentally, the word sixteen. It is not needed and it is not desirable to think of these signs as numbers, and it is very undesirable to get at the word sixteen by any process of calculation. The immediate response is most efficient and characterizes every good adder. It is obviously a matter of verbal association and no sense of further significance need clog the mind of the learner in the middle and upper grades at least.

Besides being simple, the material in adding (as in the other three simple operations) is limited. The combinations or associations required are not numerous. The combinations from one to ten are the square of ten or one hundred. If we take the digits up to nine, the square of this is 81, and this includes combinations which are the same when reversed.

Printed lists in columns were prepared, containing these combinations arranged in horizontal pairs and in random order. The children were tested by asking each by himself to add these pairs as rapidly as possible; the teacher who tested them being provided with a stop watch, and a similar list but with the answers put opposite each pair. The teacher's attention was thus free to check any pair incorrectly added and any pair in which noticeable delay occurred. These delayed or wrong combinations troubled certain

children and it was also found that often the cause of delay was that a wrong result had come to the child's mind and that he had corrected it mentally before giving his response aloud. The child was then given this new list and told that he needed to practice these particular cases. He was asked to put the right answers opposite these but to cover them with a card and to remove this immediately when he felt any hesitation, so as to read the correct answer at once. As will be shown later the children frequently use incorrect roundabout methods of adding, which even when they were accurate interfered with speed. The object of the recommendation was to get the correct and immediate association established in the child's mind by means of repetition.

The results of the test were put upon the blackboard in the form of a frequency table or in a graph showing the attainment of the children in columns whose heights were determined by the number of combinations the child could add in one minute (see Fig. 1). The analysis thus made by means of the test was presented to the children and was easily comprehended by them.

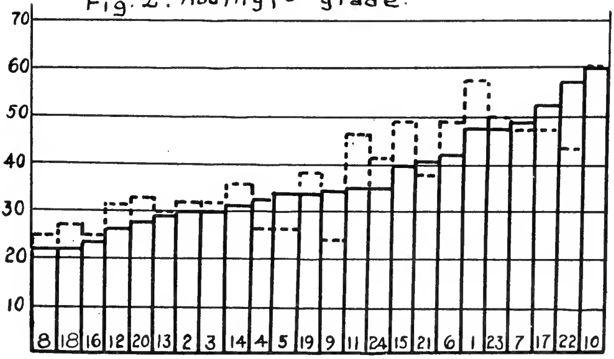
We were now in a position to ask them what should be done about it. Those who were low in adding wished to improve. Those who were high wished to know what standard should be attained. We adopted the standard of sixty combinations in a minute for the 8th grade and told the children that it was not necessary for them to practice further after they had reached this point, and that their time could be put on something else. Many children, however, practiced after they had reached this point. Never more than five minutes a day was allowed in school for practice.

The method proposed was that each child should get some other child to hear him and check his delays and inaccuracies when he practiced, just as the teacher had done in the test. When any child thought he had progressed to an appreciable extent, he came to the teacher for a more official test, and was moved up on the frequency table. The children who needed practice often came to school early in order to carry this on. At home they also did some practice, getting a relative to hold the check list for them. They were told that they would probably get better results in a shorter time if they practiced five minutes at a time and repeated this several times. They were asked not to practice more than twenty minutes a day, in order not to interfere with other studies. The carrying out of the work was left entirely in their own hands.

As will be noticed the children used the tests themselves and after the first analysis had been made, pursued methods which were understood by them and were partly their own invention, and even when suggested by the teacher were always under their own choice and control.

On January the seventh, all the children were tested by the techer with the results shown in the graph. All but three of the children had arrived at the sixty combinations a minute standard. Number 19 was out of school most of the time through illness and the Christmas holidays have intervened. Omitting the holidays this leaves about six weeks of practice. As the graph shows the average had moved up from 52 to 68. Judging from the results obtained in the 5th grade of the same school, this advance is as much as had been made in three previous years of their school attendance (see Fig. 2).

Fig. 2. Adding, 5th grade.



It will be noticed on comparison of the first 8th grade test with that of the 5th grade, that the variation is greater in the 8th grade. It would seem probable that the reason for this is somewhat as follows. In the 5th grade the children are near the point where adding (and the three other simple operations) have been emphasized as such. From this time on the work in arithmetic would assume adding to have been learned. The children would then get only the incidental practice which came from the solving of problems. In doing these many children were doubtless slow and would not get through when the rest of the class had already finished.

These pupils, would, moreover, make more mistakes in adding. As a consequence the pressure on their attention because of their difficulties in adding would prevent them giving their best attention to the nature of the problem itself. The result would be that both these children and the teacher would be disposed to think that they were poor in arithmetic and could not readily understand the problems. The children would become discouraged, would do less work, and would get less incidental practice in adding itself. This would account for the children at the lower end of the class who had not improved beyond the average of the 5th grade.

A further analysis was made in order to discover the roundabout methods used by the children. The adding list was put in a machine which exposed one pair of digits at a time and which could be regulated so as to run at the average rate at which the child was able to add. Those combinations therefore which fell below the average were thus picked out, since the child failed to give a response before the next pair was presented. When this occurred the machine was stopped and the child was asked how he had tried to add. It was found that he could give a very clear account of this. The small numbers below each individual on the graph show the kind of errors for each child as shown in the list of such errors given below. They do not show the number of such errors made in a given time. It was however easily observable that the frequency of these errors decreased in proportion to the adding ability of the child. The two who stand at the head of the frequency table (Nos. 2 and 17) were the only ones entirely free from these roundabout methods. The following list will show the kind of errors referred to.

1. (a) Doubling one number and adding or subtracting 1. Nine persons.
   Combinations affected, 7+8; 9+8; 6+5.
   (b) Doubling one number and adding or subtracting a number larger than 1.
   Combinations affected, 6+8; 8+5.
2. Using a familiar combination and adding or subtracting 1. Ten persons. Combinations affected, adding 1. 9+4 (9+3); 8+5 (8+4). Subtracting 9+5 (10+5); 9+4 (9+5).
3. Adding to make 10. Three persons.
   Combinations affected 7+9; 6+9; 7+4.

4. In combinations of 8 or 9 subtracting 2 or 1 respectively from the smaller numbers to get the last digit of the result. Three persons.
   Combinations affected, 9+6; 8+5.
5. In combinations of 9 adding 10 and subtracting 1. One person.
   Combinations affected, 8+9.
6. Adding by 2's. One person.
   Combinations affected 8+4.
7. Auditory memory, *i. e.*, had to whisper combinations before giving them. One person.
8. Finger adding, actually or by thinking of fingers. One person.
9. Learned wrong result. Five persons.

It will be noticed that the graph for the 5th grade represents the record of two tests, both of which were taken about the same time, and with no special practice between. This is for the purpose of determining the reliability of the tests. If two tests are taken on or about the same day, the relative positions of the children in these two tests ought to correspond rather closely if the test is reliable. If the children are not used to the test, there will also be a slight improvement, on the average, in the second test. But variations occur either on the part of the teacher who gives the test or on the part of the child, which will make the record of one test a little different from the record of another. If the correspondence in relative position of all the children (or correlation as it is technically called) is perfect, it is spoken of as 100%, and from this it falls off until a mere chance correspondence obtains. The amount of correlation is easily calculated and in this case was found to be 85%, therefore representing a high degree of reliability on the average for the test in adding.

As we have already said, similar work was done on the other three simple operations. On comparing the correspondence of standing of the pupils, or the correlation between the first set of tests and the last tests after systematic practice had intervened, it was found that the correlations had largely increased. The correlation between adding and subtracting, for example, before practice was 64%, and after practice 92% or higher than for two tests in adding. This means that there must have been children in the first tests who had not reached their natural place in one subject or the other and that

practice had given them a standing more nearly corresponding to their natural capacity for learning. It was found also that the correlation between silent reading and adding, although not high (15%), had increased considerably (to 30%) after practice in adding.

As suggested earlier, adding is really quite like reading. The child actually reads the figures and responds by work pronounced mentally or spoken aloud. Reading also involves habitual associations as in adding; as a result of which conceptions arise in the mind, having a mental sign, sometimes a word, not always in the text itself. One reads faster when the style is such that the words in the first part of a sentence have stimulated associations represented by words in the latter part of the same sentence. The same thing is true in passing from one sentence to another, and for that part of it the feeling of the continuity of the theme itself is caused by an expectation being aroused which is later on satisfactorily fulfilled. It is not too much to say then, that adding (and indeed mathematics more complicated than the simple processes) may be regarded as a kind of reading. In support of this we may observe that one operation in adding must take the use of three words, although these in the case of fast adders may not be fully pronounced. Some experienced adders can add about 230 combinations in a minute. This would mean about 700 short words per minute, and I have known readers who were capable of reading (silently)easy texts containing many short words at this rate. In the 8th grade graph the best adder (No. 17) adds about 90 combinations a minute. At the rate of three words per operation, this would give 270 words per minute, which is about the rate at which he reads easy texts. No. 19 is a slow reader (about 150 words a minute), No. 4, however, is a much faster reader and this probably has some connection with the fact that she improved so rapidly in her adding. If we count three words for a single operation in adding, the children would be using on the average three times sixty-eight or 204 words a minute. Their reading tests show that they read on the average 210 words a minute. The fact is, however, that many children use more than three words in each adding operation. Some of them put in the word "and" or "plus" and others both "and" and "is" or "are," all of which might better be left understood in silent adding. At the rate of five words per adding operation the words used would be on the average 68x5, or 340 per minute, a higher rate than the children actually read at. The value of this

part of the analysis for the individual child is two fold. First we may show him the advantage of using three words only. Second we can hold up as a reasonable attainment for each child the rate he has already reached in reading.

We have spoken of the child's natural capacity. This is, of course, rather an undetermined quantity. In the case of adding it would seem as if besides attention, will, or ambition, susceptibility to social stimulus, etc., that general ease or difficulty in making simple associations might be a factor in the capacity to add. To arrive at this one would need to test a considerable number of different kinds of associations, and this we did not have time to do. We selected one test, the letter-square test, often called a memory test. A card with twelve letters arranged in a square was held up before the children, for twenty seconds. This was followed by a period of sixty seconds during which the children were employed in counting aloud in order to distract their attention from the letters they had just learned. When this ended the children were asked to reproduce the letters on the card. This was repeated five times and a frequency table found of the average score. A certain power to form simple associations of letters and to remember them for a short period was thus tested.

A comparison was then made as to the correlation or correspondence in standing between this test and the two tests in adding. The correlation with the first test was 48%, with the second, 57%. Here as in the other cases mentioned, practice had increased the correlation. Those with the better powers to associate letters and remember them, had on the average made greater progress during their practice period than had the others.

The analysis has concerned itself with digits up to 9. This is of course the foundation, but ordinary adding often comes in columns, in great part not reaching up to combinations of 100. This is quite a different operation mentally from simply adding a pair of digits and requires taking into account the tens as well as the units. In order to test this a list was made consisting of numbers with two figures paired with single figures, as 29, 7; 63, 8; etc. These pairs were listed in columns and the children were tested as in the other adding test. The average was only about 20 operations a minute, and after practice of nearly two weeks had arisen only to 25. The correlations of the earlier of these adding tests with the first adding test, using single digits, was 75%. After practice the

correlation had risen to 80%. It is obvious that much more practice was needed for the latter form of adding.

A consideration of these facts would seem to justify the conclusion that considerable correlation even when subjects are not closely related would indicate a general state of efficiency in the work of the school. It would show that although the pupils differed among one another, yet that each had approximately made the progress which his natural powers justified and had not been neglected in any particular subject on one hand, or driven to excessive effort on the other.

Problem Work in Arithmetic

However important the four simple operations are, they are of course only subordinate to the problem work. In order to make this real, to connect it with the children's interest and bring it into contact with their environment, we asked the children to tell us some of the things they had wondered about, and that perhaps some calculation would solve. One girl had wondered for years how many feathers were in a pillow. Another pupil had wondered how much ensilage his father's silo contained, etc. We considered how these problems could be solved. A certain number of feathers were counted and were then taken to the drug store to be weighed. The pillow full of feathers was also weighed, and from these data the number of feathers approximately obtained. The cubic capacity of a certain weight of ensilage was found, with the cubic capacity of the silo, and from these data the problem solved. In these cases the children discussed the problem in order to find what to do with it. This was done in committees with the teacher helping as the pupils required.

After the results were obtained they were presented to the class as a whole by the individuals or committees who had worked them out. The children began to work upon arithmetic as a real game which could be applied to things they came in contact with. They began to take an active attitude of investigation instead of passively receiving methods from the teacher which they were not able to apply.

Silent Reading

There is no subject in the school curriculum more important than silent reading. This is true not only for the elementary school, but for the high school and college also. In these latter institutions,

many pupils fail in their work because of their inefficiency as readers. Reading is not specifically taught in the high school; nevertheless the pupils learn to read during their high-school course. The average rate of reading for pupils at the end of the 8th grade, using easy tests is 210 words a minute. At the end of the high-school course, pupils on the average read the same text at the rate of 300 words a minute and get a proportionately larger number of ideas from what they read; while college seniors have an efficiency of 330 words a minute.

For the future citizen the capacity to read easily is more important than any other single subject he learns in school. The kind of material he reads when he leaves school is determined almost wholly by his efficiency as a reader. It is not so much the so-called sensationalism of many newspapers that leads to their wide circulation as the fact that they appeal to a class of inefficient readers. A reader who cannot read easy text faster than 100 or 150 words a minute (and there are many such even among high school graduates) will prefer newspapers where the sentences are short, the vocabulary limited, and the ideas trite and commonplace. The paper will then be largely filled with immense headlines, photographs, a kind of humorous picture writing, cheap stories and editorials in large primary grade print. If he reads the ordinary newspaper at all, such a reader will usually omit the editorials, which even for efficient high school and college students are not read faster on the average than about 250 words a minute.

To meet this situation the work done under the name of reading in the most of 8th grades is quite inadequate. In the more backward districts a book called a reader is usually provided, containing perhaps 60,000 words. If the children should or could read this at a normal rate, it ought to be finished in five or six hours. This is not the way however, the book is treated. Portions of it are too difficult to read easily and can only be stumbled over with little comprehension of the meaning. If any portion should prove to be interesting in spite of its difficulty, it is often only a small section, perhaps only a few paragraphs taken from some larger text. In the less backward districts supplementary texts are provided, but these, as well as the readers, are too often used not for the purpose of improving the children in silent reading or in getting the thought rapidly and accurately from the printed page, but for the purpose of oral reading for literary discussion. Quite often a text much

beyond the children's power to read is taken. Stanzas or paragraphs are discussed separately and reduced by the teacher to language the children understand. The pupils get the story from the teacher rather than from the text and although they may be interested in this, the process is slow and does not give them the independence of good readers.

An example will illustrate the traditional method. Before we began the demonstration work in reading, the 8th grade was occupied with the *Lady of the Lake*. The class studied or read silently one or two stanzas, after which some pupil was called on to read aloud. On one occasion when I happened to be present, a boy read:

"No thought of fear, no thought of rest
A sausaged the storm in Roderick's breast."

To the eye the right word "assuaged" looks a good deal like "a sausaged" but to the ear the difference is very marked, and the class burst out laughing, as did the reader himself. However, this was not because they really knew the word assuaged, but because they knew that sausages, at least, had nothing to do with the case. On being questioned as to the meaning, someone in the class said the word meant stopped. The teacher helped them to the word calmed, which was, of course, better. In both of these cases the minds must have worked by inquiring what is the word and the idea that is to be associated with the word and idea of storm, thus in both cases translating Sir Walter Scott's language into something more comprehensible. I asked the children at this point why Sir Walter Scott did not say calmed instead of assuaged. Some said on account of the metre, but thought the metre would have been preserved if he had said, "calmed down;" and they would have liked this better. I asked them then, "why didn't he say calmed down?" Some said, "it was because he was a very learned man and he wanted to show this off." I asked them if they liked that kind of man. They replied that they did not. Some then suggested that he put this word in so that we could learn it in school, that there were schools in those days and that the *Lady of the Lake* was probably written as a text for them. I asked them (since they thought that was the reason) whether they had now learned the word and whether they would be able to use it again; but they could think of no situation where they would ever have the courage to use this word.

I then told them the word was used in connection with thirst to mean the satisfying or partial satisfying of thirst, and asked them what could be regarded as corresponding to thirst in this case. Someone said at once that the storm referred to was the storm of anger and that this was like a thirsty beast. They began to admire Sir Walter Scott's use of the word, but still thought it unlikely they would ever use the word themselves.

If there were only a few words of this character in the text, it might be suitable enough to develop their meaning at length, but we found that there were some thirty words in every canto which the children either did not know at all or knew very imperfectly indeed. It is very important for the purpose of reading to add words to the children's vocabulary, but the words added should bear a relationship to the words already in stock both in kind and in quantity and should as much as possible consist of words capable of being used frequently by the children. To do otherwise is to miss the mental region or budding point at which the children can grow most rapidly and efficiently. It did not seem to us that the *Lady of the Lake* was the best material for this particular 8th grade, if we wished to give the kind of practice in silent reading that would be sufficiently independent and self satisfying to be continued when the pressure of the school and the constant aid of the teacher would be removed.

As a preliminary to further analysis of this problem we proceeded to test the children's silent reading capacity.

It is comparatively easy to get a test that will determine the average reading capacity of a whole class in comparison with other classes of the same or different grade, but it is a more difficult matter to determine reliably each child's separate standing. We selected reading matter from two sources. Narrative and descriptive material from *Irving's Sketch Book* and more abstract newspaper material from *Current Events*, of which we had copies every week for every child. The children were asked to read for one minute and to write out immediately afterwards a full account of what they had read. The number of words read in the text gave us the score for speed in words per second, and the number of words written in the reproduction, after repitetions and errors had been crossed out, was taken as a measure of the children's comprehension. In some preliminary tests we had found that counting the details or ideas of the reproduction gave us about the same results as counting the

number of words written (as Starch has already found). We, therefore, followed the latter method as it was simpler and occupied less time.

The correlation (or correspondence in standing) between speed and comprehension in a single test lasting one minute was not very high. For the material we used the average correlation of four such tests was found to be only 52%, ranging from 42% to 78%. The latter correlation was found in a case where the reading material (from the *Sketch Book*) was full of action, easy to grasp, and therefore easy to remember. Figure 3 shows the correlation table for this case. The correlation evidently means that on the average the child who reads quickly has passed over more ideas and therefore has more ideas to reproduce than the child who reads slowly; or to put it from the other side, the child who comprehends a given idea or detail in the text sooner than his neighbor, will pass on to the next idea or detail more quickly and thus will be reading at a more rapid rate.

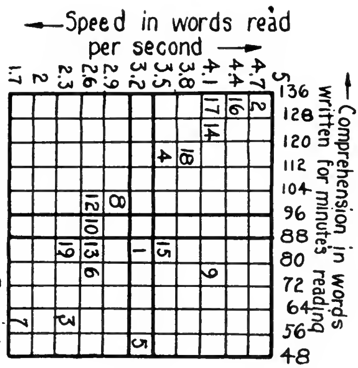


Fig. 3. Silent Reading

There are many factors however which reduce this correspondence between speed and comprehension, especially when the latter is measured by a written reproduction. It takes some ten or fifteen minutes to write out what has been read in a minute. During this time a good deal that was in the mind at the time of reading, has lapsed. If it were read again the portion might be recognized, but it is not in a position to be voluntarily recollected and written out. Sometimes whole sections will thus drop from the child's mind, and

although he knows there is something further to be recovered, he is not able to get a clue. Occasionally after he has written his paper, perhaps when he is leaving school, inhibitions are removed and the passage will suddenly flash upon him. In such cases as these the child's record for speed will stand relatively higher than his score for comprehension, and if this happens among a number of children not equally but in varying degrees, the correlation or average correspondence between speed and comprehension will be proportionately reduced.

But it is not lack of memory only which reduces the correlation. There are different degrees of comprehension even at the moment of reading. How much comprehension of a given portion is necessary in order to release the reader's attention for the next portion? With some readers, very little. I once knew an old man suffering from senile dementia who was very fond of reading and would spend hours with a book in his hand (mentally pronouncing the words, since one could often see movements of the lips and throat,) but who could give absolutely no account of what he had read even when the passage contained but two or three sentences. There must have been practically no associations of meaning aroused by the words he pronounced, and yet the practice of a well established motor-auditory habit succeeded in giving him pleasure and satisfaction—perhaps something like that of twiddling one's thumbs. This is of course an extreme; but there are a good many readers who approach it, if not for all parts of what they read, at least for certain portions. In such cases speed of reading may be rather rapid, but since there is little to reproduce the score for comprehension will not correspond; and when this happens for different pupils in varying degree, the correlation will necessarily fall.

The opposite also obtains where speed is slow and comprehension relatively large. Here there may be mechanical habits such as irregular and too frequent eye movements, or the habit of too complete pronunciation, both of which lower the speed while the mind makes use of the extra time thus forced on it in attention to details. Or, again, the comprehension may be exuberant and imaginative in the sense of revelling in details or situations of a more or less subjective character, and not strictly necessary to the understanding of the text. In such cases the reading rate is somewhat slow while the reproduction is copious if not always accurate.

The best reader falls into neither of these extremes. His attention does not dwell too long on any detail. He builds out the meaning of the text and organizes it into wholes. Meanwhile the mechanical process of moving the eye goes on with great regularity while the mental pronunciation follows behind somewhat elastically, at times being two or three words behind the eye and at other times ten or a dozen according to the difficulty of comprehension.

As will be seen from the analysis above given, the measure for speed on the average, indirectly and approximately, measures the comprehension and vice versa. The average of the two ought therefore to give a fair measure for reading capacity as a whole. In order to get numbers that are of the same dimensions and will therefore average, it is necessary to reduce the score for speed to comprehension, or the score for comprehension to speed. On the correlation table No. 9, for example, stands at 4.3 in speed and at 77 in comprehension. If we wish to translate this latter score into speed, we should take the corresponding position in speed. Since No. 9 is in the fourth column from the bottom in comprehension, we should therefore substitute the score of the fourth column from the bottom in speed. This is 2.7. The average of this with the average for speed actually obtained (4.3) is 3.5. The single score of No. 9 for reading efficiency on this particular test is therefore 3.5 words per second in terms of speed.

It will be seen from the correlation table that for this test the children write as their proof of comprehension about half as many words. as they read (accurately 45%). This is for one minute's reading. Starch used half a minute's reading and found for the 8th grade a comprehension of 49 words in half a minute and a rate of four words per second. The number of words written was therefore 41% of what had been read. The total efficiency, however, is greater, but we tested the children later with Starch's test and found that their average was equal in this case to that found by Starch. It would seem therefore that reading for one-half minute gives a higher record but that this comes in from the side of speed rather from that of comprehension. As is well known the first part of a piece is better remembered than the last part. Merely to cut off the last part, however, does not seem to have the effect of giving a larger proportional reproduction. The lower proportional reproduction in the case of the half minute's reading may be due to

the shortness of the piece which would prevent the children's getting enough of the story to build out a thoroughly coherent picture or general conception. Many touches required to illuminate the understanding of the earlier part of the reading come in the latter part. The earlier part, therefore, has the advantage of this emphasis which amounts to a kind of repitition in memory and this would be one cause for the earlier portion being remembered better.

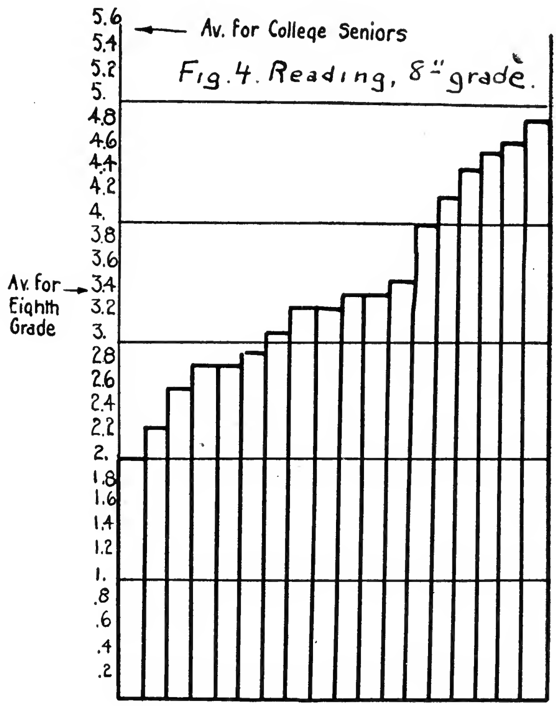
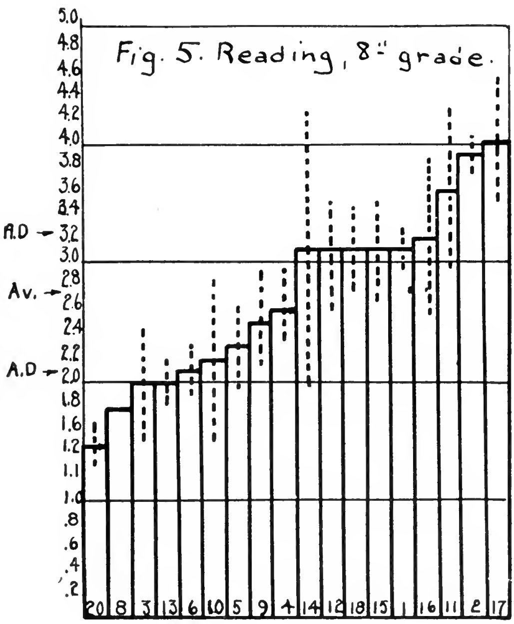


Figure 4 will show the general efficiency of each pupil in a single score made by combining the score for speed and that for comprehension. The average is 3.5 in words per second. The average reading efficiency, using the same test for college seniors is 5.5 in words per second. Another test using the same kind of material (from the *Sketch Book*) gives an average for the 8th grade of 3.6. The correlation between these two tests is 85%, thus indicating their reliability.

These tests, however, are from the same kind of material and will favor some pupils at the expense of others. This is shown by the fact that when a test from *Current Events* was correlated with one from the *Sketch Book,* we found the correlation fell to 40%—as compared with 85% with a pair from the *Sketch Book,* and 75% with a pair from *Current Events.* This means that there were some children who could read the *Sketch Book* easily as compared with *Current Events* and vice versa (allowing of course for the greater general difficulty of *Current Events*). When, however, two tests from the same kind of easy material (*Sketch Book*) were taken, the individual scores averaged, and this score compared with another score made in the same way from the same kind of material (*Sketch Book*) it was found that the correlation rose to 90%. An average of two tests from easy material is thus a very reliable individual test for reading material of this kind, but should be combined with tests from harder material if we wish to get a measure of general reading ability.



To arrive at this we took two tests from the *Sketch Book* and two from *Current Events.* Figure 5 shows the results. The individual average deviations for the four tests are shown by the vertical

dotted lines running through each score. The average for these taken together is .43 (average intra-variation). Compare this with the average deviation for the score as a whole (.62 inter-variation) and one realizes what a large extent of the total range may be covered by one pupil. No. 14, for example, in one minute's reading, may nearly be at the top of the list and in another nearly at the bottom. Great variability of this kind, however, throws an interesting sidelight on the general makeup of the pupil. This variability is characteristic of a large part of this pupil's work at his present stage of development. He is enthusiastic about some things, bored and inattentive with others. His mind is filled with a certain class of images, especially the romantic and exciting kind, while the more abstract and sober ideas are scarce. His high scores in reading naturally come from the *Sketch Book;* his low scores from *Current Events.*

A much simpler way of testing the children's reading capacity would be attained if the need for the record for comprehension could be eliminated. In order to accomplish this end the writer had a machine constructed for reading which exposed succeeding sections of a newspaper line in such widths as correspond to the widths taken in at a single glance by an efficient reader. On a newspaper line this amounts to three or four stops per line. The machine is so geared that its speed may be changed without interruption to suit any rate of reading. The slow reader may thus make two or three stops for each aperture if the machine is running slow enough, but if its rate is increased he is forced to move on and will probably decrease the number of stops. If the rate is too fast he will not be able to make sense of what he reads, but there is a point in the speed of the machine where he sees all the words, pronounces them mentally, and is able to understand their meaning The most rapid rate at which he can do this is of course his rate of reading for the material in hand. The number of lines per minute at which the machine is running may be read off on a scale and from this the number of words per second easily calculated. As a farther control the machine may be stopped at any moment and the reader asked to say what he has read. Practically this is not necessary as the reader soon complains if the machine is going too fast for him to follow. In any case there is no need for a special score in comprehension.

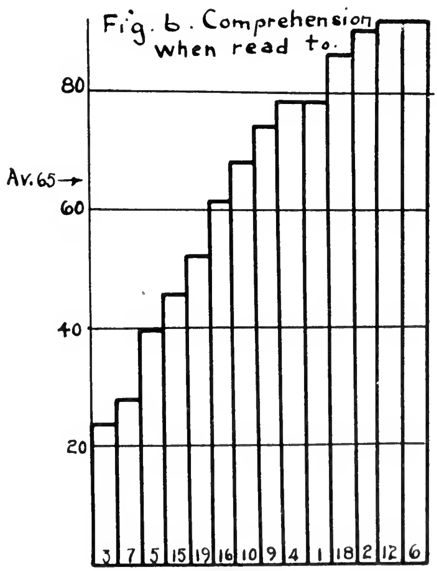
Apart from its facility in testing individuals, the reading machine is of value in improving the habits of slow readers in as far as these are due to irregular and too frequent eye movements or to complete and painstaking mental pronunciation. By gradually gearing up the machine with readers of this class, a very great increase in their rate of reading actually results, the average going up with easy texts to 4.5 words per second. At such a rate the attention must be on the alert at every moment. There can be no idling and returning to words already seen. The rate of eye movements must be regular and the tendency must be to approximate the eye movements of a good reader.

The machine is useful moreover in analyzing the difficulties of readers in as far as these are due to lack of comprehension. If the machine is running a little faster than is comfortable for the reader, phrases occur which although pronounced are not fully understood. An instance will make this plain. A pupil had read a sentence containing a statement "Our soldiers had gone abroad on a mission of patriotism." The machine was stopped at the end of this sentence and the reader asked to say what she had read. She gave the meaning of the sentence but left out the idea expressed by the words "mission of patriotism." "But there is something you have left out," I replied. "Yes," she said "there was something else I didn't get." The machine was returned to the beginning of the sentence, and she read it again at the same rate as before. But it was not until she had read it for the third time that the meaning of the phrase dawned upon her. I asked her why she had not understood it and whether she had known the meaning of the words "mission" and "patriotism." Oh, yes, mission has to do with missionaries. We send money to them in church, and patriotism has to do with Red Cross drives, the flag and love of one's country, but I didn't see how soldiers could have anything to do with missions."

This is a typical instance. The difficulty does not consist in the vocabulary as ordinarily understood. It consists in the unusual combination of words, each of which may be known but out of which the imagination must construct a meaning somewhat new to the reader. After this association is once learned it of course becomes a part of the habitual memorized equipment of the reader, and the meaning of the phrase is obvious at once. It is the breadth of associations that words have with varieties of different situations that makes the easily comprehending and rapid reader. Learning a single dictionary meaning therefore is not sufficient.

Reading Aloud

Comprehension of material which the pupil reads for himself leads us to inquire as to his comprehension when he is spoken to or when he is read to by someone else. In order to investigate this, two easy pieces were read to the class from the *Sketch Book*, each taking one minute to read. One piece was read at the rate of 143 words per minute (2.3 words per second) and the other at the rate of 157 words a minute (2.6 per second). This is of course a slow rate, but there are some in the class who read for themselves at this rate. Moreover, oral reading is slower than silent reading since the words must be fully articulated. Each of these readings was reproduced in writing by the class and the scores averaged. The results are shown in Figure 6. The average is only 65 while the average for self reading for the same sort of material is 94. This is due evidently to less ground being covered since the average rate at which the pieces were read (2.4 per second) corresponds to a reproduction of 70 words for pupils who read at this rate (see Figure 3). It is evident that the most of the class have been sacrificed to the slow rate at which the pieces were read, and they have not made up for this by remembering a greater number of details. No. 2, for example, writes only 84 words, while when she reads for herself she writes 134.



Reading aloud is a capacity which has for its principal measure of success the comprehension of what is read by those who hear. No elocutionary antics or correctness of position will make up for lack in this respect. Distinct articulation with a speed suitable for the hearers is the principal factor. In this connection it is important to notice that the rapid silent reader does not make a good oral reader without special training. In a previous investigation on normal school students (Boston) I found that there was a negative correlation of 75% between efficiency in silent reading and articulation as measured by the teacher of reading. Those who are best in silent reading, on the average, are those who do not pronounce the words fully when reading to themselves. This habit evidently goes over into oral reading. There is therefore great need for special training in this branch of the subject.

The test for comprehension during oral reading is significant not only for this exercise, but it throws a light on all comprehension of spoken discourse and therefore makes reading of the utmost significance for every subject in the school. When the teacher is speaking, some pupils are comprehending what she says much more rapidly and perfectly than others. Almost instinctively she is led to speak either at such a rate as to accommodate the slowest in the class or to use language so denuded of associations and infantile in structure as to be comprehended by her worst pupils. It is the latter alternative she usually follows, and thereby sacrifices the remainder of the class. Merely to remove a few of the worst pupils is no cure. The gradations of the pupil's comprehending capacity, as the tests show, are almost continuous. When one or two of the worst pupils are eliminated by failure in promotion or by transfer to special classes, this will raise the quality of the teacher's diction but a very few degrees.

Methods in Reading

As we pointed out in the section on arithmetic, analysis of the situation is a necessary preliminary to successful methods. The analysis of the situation in reading just made was necessary in order to show what a child is actually doing when he reads and what constitutes a reliable test for this function.

It is of course out of the question to put all of this analysis before the children, or even before the average teacher. The children, however, can learn to make the tests in reading themselves, and the more official tests made by the teacher can be placed before them.

These scores will themselves provide a considerable amount of analysis which the children can understand perfectly and which they are willing to make a basis for their practice.

After the first tests were presented to the 8th grade demonstration class, we consulted with them as to the best way of improving their reading capacity. As a result of mutual suggestions, we agreed that each child should read at home (daily) a selection from something he was interested in and was perhaps already reading, and give an account of this orally to his classmates. He was asked to time his reading and report this also. How much can be told next day of something read once and taking ten minutes to read? In order to give a proper and intimate audience, and in order that each child would have an opportunity to give an account every day, the class was divided into groups containing four or five pupils each. The pupil after he had given his account, asked the others what images or ideas they had gotten and emphasis was placed on the differences of these individual pupils. The hearers then asked questions as they wished and often brought out parts of the account which had not been given at first, a practice which tended to make this pupil's account better the next time. Criticisms were also offered as to the manner of speech. Too many hesitations—too many "and's"—couldn't hear him—didn't open his mouth enough—bad grammar—were some of the more frequent criticisms made. The group then decided which account had proved the most interesting and this was recommended to be given to the class as a whole.

After the pupils had made some progress, it was suggested that they bring the book or magazine from which they had made their selection, to school, and where suitable, read the actual words of the text, putting these into the body of their spoken reproduction.

Reading from *Current Events* was taken up by having pupils read silently for about a minute and then single pupils would be selected to give an oral reproduction. Time was not allowed for re-reading or for slower pupils to read as much as the faster ones.

Besides this, many pupils began to read somewhat harder material, such as was contained in the editorials of the daily papers. They noted any phrase with which they were not familiar and which delayed their reading. They wrote these phrases down, or sometimes clipped the line containing them out of the newspaper and thus made lists of them in a note book given them for this purpose. They brought these phrases into their groups and if members of the group

did not know them they had the privilege of adding them to their collection. The following by No. 17 is a small part of such a collection.

radical theorist
grovelling terms
peace without annexations or indemnities
vehement, tumultuous emotions
clarify the illusions
imperceptibly sweeten life
possessing the originality
unsightly and unnecessary projections and appendages
muddle-headed ignoramus
congenial companion
by their consituents
garrulous exultant throng

Another from a less efficient reader, begins as follows:

practical conclusions
general principles
Prussian oppression
in substance
radical theorists
enigmas of the war
future exigencies
commercial aspirations
congenial companions
sentiment of the country
it is incomprehensible
ingredients of clay
every one was surfeited

There is a striking difference in the character of these lists. Probably no phrase in the last list would have given any trouble to the collector of the first list, and the most of the phrases in the first, would have been unsuitable for learning by the second pupil until she had mastered commoner phrases such as she actually collected. Each pupil is thus learning individually at his own level, but what one learns is also communicated to the others and shared by them in as far as their interest and ability warrant. They are free to put phrases collected by others into their list or to leave them out, as they see fit. They were recommended not to put phrases into their lists which they did not feel they could find some opportunity to use either in conversation or in their compositions. They frequently reported some of their attempts to use their newly discov-

ered phrases at home. One boy said to his father, "What do you think of the proposed ratification of the prohibition bill?" His father said, "What's that?" On its being repeated, the father said, "You may as well talk German at once, and have done with it." Since the boy told his story with good humor, it is likely that the reaction of his father rather stimulated him than the reverse. He might have felt that he had scored one on his father, or he might have seen a lurking admiration in his father's seemingly unsympathetic answer. A girl reported that she had asked her mother to "clarify her ideas" on some matter. The family twitted her all the evening, asking her if she could clarify their ideas on this subject and on that, but it is likely that the process had helped rather than hindered.

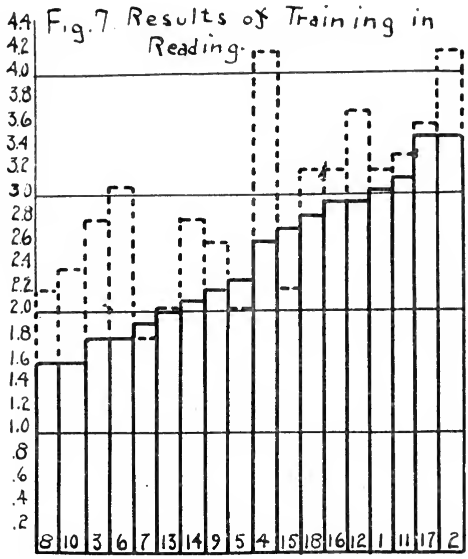
A previously arranged course of study in which a list of phrases selected by the teacher would be given to the class and learned by all, could never be so well adapted to their individual needs. This is the method of autocracy too prevalent in our schools. I think it is not too much to say that the more efficient method followed by the demonstration class, while individual, was also co-operative, and was thus truly social, democratic and American.

Results in Reading

Work of the kind described was begun in the latter part of January and the beginning of February. Tests were taken at this time and about three months afterwards other tests were made as nearly like the first as possible. Figure 7 shows the difference. Each score represents the average of two tests one minute in length from *Current Events.* The average of the first test (score combining speed and comprehension) is 2.4; that of the second, three words per second. The correlation of the two tests is 85% showing the class had progressed pretty evenly. The average deviations (inter-variation) are 5.4 and 5.6 respectively. These overlap somewhat but not enough to weaken very much the independent status of each average.

The progress is not so striking in reference to the total achievement as that already shown in the case of arithmetic, but the reason for this is that the total volume of associations to be learned in the case of reading are hundreds of times more numerous than in the case of adding. As we found in a vocabulary test, the average reading vocabulary was about 21,000 words in a test made from a

large dictionary, in which the average of college seniors was 42,000 words. When one adds to this vocabulary the many associates with each word, we have a number of mental operations compared with which the 81 associations to be learned in adding represent a very small task indeed. The proportion therefore that the children will learn in three months' time cannot be expected to be very large. Perhaps at the rate of a little over 2000 words per year.



As we have already seen, vocabulary is only one factor in reading. This is shown by the fact that when two tests in vocabulary were averaged, of which the correlation was 71%, and this score correlated with the score in reading, made also from an average of two tests, (correlation 85%) the correlation between vocabulary and reading amounted to only 52%. This shows that while vocabulary is distinctly a factor, eye-movements, habits of mental pronunciation, familiarity with phrases, interest in the content, imaginative construction must also play a large part in reading efficiency.

As the figures show, however, the progress is noticeable, amounting to a difference of .6 words per second or 36 words a minute. It is as much, in fact, as Starch found on the basis of a large number of schools throughout the country for the average growth of a whole

year between the 7th and 8th grades. That this progress was made during a practice period of three months must be attributed to some superiority in the process at work.

Composition, Spelling, Writing, and Grammar

A good part of the work the children did under the head of reading can also be classified under composition. The extension of the children's vocabulary and phrases and their practice of making reproductions both written and oral, certainly belong here, as well as under reading. But this kind of work is largely passive and is concerned with the ideas of other people. An opportunity for something more active and as creative as possible is also required.

Composition as it is often conducted in schools is not regarded as primarily the art of written communication. Children are not usually asked to write to anyone in particular unless perhaps the teacher; and in this case when the children are as old as those in the 8th grade, they frequently see through the illusion and realize that they are actually writing for the waste paper basket. When a business man writes a letter, he has some object of a social character in view. He is writing to someone and he is expecting to cause some change, however small, in that person's mind. In harmony with this purpose he selects his ideas and chooses his language. Lack of clearness, of force, deficiencies of spelling and grammar, are avoided because of their effect on the mind of his reader. He is, moreover, able to gauge the success of his effort by the response which he obtains. All of these motives for efficient and interesting writing are usually absent in the average school.

After testing the children by the Harvard-Newton scale, and finding that no composition was better than the fourth class of this scale (which runs from 1 to 6) we suggested to the children that they use the same method they had been using in reading, that is, the children wrote compositions on subjects of their own choosing which were then divided into sets of four or five and handed to groups or committees of the same number. Each member of a committee then read each paper handed to his group (none of which were his own) and arranged the various papers in order of merit, according to his judgment. When every paper had been read by each child, the committee began to discuss the merits and demerits of the composition and to explain why each had ranked the set of papers as he had. If in this discussion any child saw good reason to change

his ranking of the papers, he was at liberty to do so. The average standing of the different compositions was then taken by adding together the rank given by each individual reader and dividing by the number. This was reported to the teacher and the class and the best composition recommended to be read by the writer himself to the class as a whole. While the discussion as to the comparative merits of the papers was going on, the teacher passed from group to group, taking part as she saw fit. She was frequently appealed to on various points of structure, grammar, usage, etc., and took the opportunity thus afforded to give the children the instruction desired. The different sets of papers were then taken to the College and a group of five or six seniors read and ranked them in the same way the children had done, but without knowing previously what the children's ranking had been. On no occasion was the average ranking on the part of the students different from that already found by the pupils of the 8th grade.

We asked the children whether they thought any of their parents or other relatives would care to read some of these sets of papers and rank them as the college students had done. The children thought a letter from the teacher to the parents would help matters. It was suggested that the class compose a suitable letter, which, after passing everybody's criticism, would be signed by the teacher and sent with a set of compositions to the various homes. This was done, not so much to get a still more objective standardization of the papers, but to interest the parents in the work of the school.

In order to test the power of 8th grade children to judge compositions, not their own but written by other 8th grade children, we had copies made of a descriptive set in the Harvard-Newton scale. The children were asked to grade these as they had been doing already with the compositions of their own class. The judgments of one group are shown in Table I. The other groups gave similar results. The order given to these papers by the eighty 8th grade teachers, who judged them first (although their variation was large) is the same as the order found by the children, except that the "Lake at Sunrise," and "The Mansion" are in reverse order. These compositions are, however, practically paired by the children. My experience with adult judges, moreover, shows that these two compositions are frequently interchanged and are actually harder to discriminate than any of the others. In a previous investigation made on the 8th grade of the Everett School in Boston, I found

further evidence of the capacity of the 8th grade children to judge these compositions. Of 37 pupils in this grade, the average judgment was for "The Storm in the Fishing Village," 1.84; "The Grandmother," 2.03; "A Mansion," 3.13; "The Lake at Sunrise," 4.05; "The Lighthouse," 4.50; "A Scene on the Prairie," 5.35.

TABLE 1. PUPIL'S JUDGMENTS OF COMPOSITIONS

Correct order (Teachers)	Pupils judging	3	5	2	20	9	19	Av..
5	The Lighthouse	4	3	5	5	5	5=27	4.5
2	The Grandmother	2	2	3	1	2	3=13	2.2
3	A Mansion	3	5	4	2	3	4=21	3.5
1	Storm in the Fishing Village	1	1	1	4	1	1= 9	1.5
6	A Scene on the Prairie	6	6	6	6	6	6=36	6.0
4	The Lake at Sunrise	5	4	2	3	4	2=20	3.3

It would seem from these facts that a random group of five or six 8th grade children would in the most of cases arrive at the same, or nearly the same, average judgment as the teachers of this grade.

When the next compositions were judged, the sets of papers and groups were differently arranged, so as to give each pupil a chance of being compared with other pupils in the class. A pupil who had been judged by a committee in one case as first, might be judged, when compared with other writers as second or third, even when his paper was equally good. In estimating the pupil's real capacity these conditions were naturally held in mind, both by the writer and his judges.

There is something to be said in favor of judging the pupil in comparison with a small number of his classmates rather than with all, even if this latter could be done successfully either by the teacher or the pupils; because although the pupil receives some honor or the reverse, there is no reason for him to be either puffed up or discouraged. It will only be after a large number of such tests, that he and his comrades may estimate finally his general ability as a writer. Meanwhile there is plenty of opportunity for him to try various ways by which he may please and impress his readers.

Work of this kind combined very well with the work in reading. I was present once when the best compositions from different groups were being read to the class as a whole. Approval or the reverse was expressed by the members of the class. "I like your saying stately maple, you might have said, "stately maple tree," or "Did you get that phrase 'the mellow light of the moon,' recently for your collection?" "Yes," answered the writer.

As has been said, at the beginning of this work we could judge no composition higher than the fourth position on the Harvard-Newton scale. After four months' work, four writers at least had moved up to the third position, while only two or three were still left behind in the fifth. Lack of space prevents us giving samples of these compositions.

While this work was going on, we tested the spelling, the penmanship and the formal grammar. In the latter subject the children stood high, although there was no particular evidence of it either in their speech or conversation. In spelling (Ayres' scale) they were found to be about a year behind their grade, and in penmanship, about on the average. We did nothing special to improve these subjects, except in as far as they entered into the composition. The teacher meanwhile conducted recitations in these branches as formerly.

Neither had we time to make anything more than a beginning in history and geography, subjects which promise large returns, both from the standpoint of increased efficiency and co-operative democratic methods. The manual training, drawing and music were taught by special teachers and were unrelated to the rest of the school work. Nothing was done in these subjects, either, towards increasing their efficiency. We look forward to the work of another year to continue what we are now doing and to add some other subjects to our field of demonstration.

The work of the year was made possible and enjoyable by the loyal and intelligent co-operation of Miss Marion Chase, the teacher in charge, and every courtesy was extended by Mr. Whittemore, the superintendent. It is expected that this demonstration class will be made a station for observation first for the teachers of the village and secondly, although indirectly, for other teachers throughthe country.

THE EFFECT OF EXEMPTING PUPILS PROFICIENT IN HANDWRITING

CYRUS D. MEAD
University of California

The pupils of an ordinary fifth or sixth grade class will write with varying degrees of legibility. There will probably be some pupils as low as quality 6 or 7 and some as high as quality 13 or 14 by the Thorndike Scale, this being nearly the complete range on the Ayres Scale. It has been found that, roughly, one third of the pupils of any one grade, from the first to the eighth inclusive, will reach or exceed the median quality of the next higher grade; and that one fourth of the pupils will reach or exceed the median quality of the second grade above it. For any other subject than writing, this condition might become the basis for a differing gradation. It is much less a problem in writing in that class instruction and practice in this subject may include a wider range of abilities without impairing the efficiency of such instruction. But why should individual pupils who have acquired the writing abilities of so called "standards" above their grade and show a reasonable speed, be required to take formal instruction in writing so long as they maintain such a quality? The very suggestion however, of exempting such pupils from "regular writing periods" will, almost without exception, be met with indignation by a supervisor or teacher of writing. It is as if some sacred custom had been violated beyond repair. If you protest that you have no pre-conceived notions regarding the matter but that you wish to get experimental evidence upon which to base some notions as to the effect, you are apt to be told that handwriting is too important an accomplishment to be experimented with. Yet such an attitude of mind prevailed in science from Aristotle to Galileo, and history records many of the intervening centuries as the "Dark Age."

To find whether or not the posting of a handwriting scale in a class-room was any incentive to improvement in writing and whether

[1]With the cooperation of class teachers from day to day, the following members of my Educational Measurements Seminar in the College for Teachers of the University of Cincinnati conducted the preliminary and final, and also the intervening monthly progress tests:

C. M Howe in the Fairview School; John W. Snyder in the Madisonville School; and John F. Arundel in the Lincoln School, Cincinnati.

or not exempting especially proficient pupils was detrimental to them, and to class progress in writing in general, an experiment was carried out over the greater part of the school year of 1917-18 with three fifth and two sixth grades, in all, 203 pupils.[1] The problem and plan stated briefly were as follows:

A CONTROLLED TEST IN HANDWRITING

Problem: 1—Does the quality of handwriting in a class suffer by exempting especially proficient pupils from formal writing periods?

2—Does the posting of a handwriting scale effect improvement in writing?

Plan Three classes each of fourth, fifth, and sixth grade pupils selected and divided into three groups as follows:

Group I—One fourth, one fifth, one sixth grade.

1. Sixty minutes weekly for formal writing exercises.
2. Gettysburg Scale posted and explained.
3. Pupils encouraged in using same for comparison of work from time to time.
4. Monthly tests for scaling.
5. Those pupils who reach speed and quality of one grade above exempted from practice at formal writing periods, this dependent however, on their maintaining same in each monthly test.

Group II—One fourth, one fifth, one sixth grade.

1. Same
2. Same
3. Same
4. Same

Group III—One fourth, one fifth, one sixth grade.

1. Same

(A preliminary and final test)

The original plan as seen above contemplated three groups of classes, a fourth, fifth and sixth grade in each of the three groups and three different schools. Due to various difficulties, the final data could be completed only for the three fifth and two sixth grades. These classes were of the regular school organization and from the nature of things not necessarily of equal abilities within their re-

spective grades. They also had different teachers. These are qualifying factors for which the reader must allow. It will be seen however from the scores below that Groups I and II, that is, the fifth grades of these groups, in the initial tests, were of practically the same average and median abilities in quality of writing. It is in these groups that the chief interest centered. Groups I and II were "German" sections. Some time during the day was given over to instruction in German. (This has since been done away with). These groups had the full benefit of the posting and featuring of the Ayres Gettysburg Handwriting Scale and of monthly tests therewith. The chief factor of difference however was that the pupils of Group I had the added incentive of exemption of pupils each month from all formal writing periods when such pupils showed sixth grade ability, and held it, in both speed and quality.[2] This gave them the three twenty minute periods weekly for other more needed work.

Group III (the fifth grade) was a regular "English" class with no instruction in German. It was probably a slightly inferior group in that it included some failures from the classes taking German. The average and median initial scores in writing show them to be about one point lower on the Ayres Scale, although more rapid writers. This group had no knowledge and made no use of the scale in the class-room. It was given only the initial November and final May tests. Condition one, however, was the same, *i. e.* three twenty minute formal writing periods weekly. The purposes of the controlled tests were, again, to obtain some objective data for altentative answer to two questions: First, did the incentive used of posting and explaining a handwriting scale in the school room prove to be an effective agent in improving handwriting as shown by comparison of Groups I and II with Group III; second, did the described exemption from all class practice of proficient pupils lower or raise the standards of their groups, as shown by comparing Groups I and II with each other? The results with the fifth grade classes will be taken up first. When referring to groups, this grade only is meant. For purposes of illustration, the records of one class are given in Table I. The intervening monthly scores are omitted but the leaded lines mean the months over which the pupil was required to take formal instruction and practice in writing. A space means that the pupil was exempt. All tests were given during the first week of each month.

[2]By sixth grade ability is meant the average number of letters written per minute and the average quality of writing "commonly found in grades".

The following averages are taken from the Ayres Gettysburg Scale:

Grade	Speed	Quality
2	31	38
3	44	42
4	55	46
5	64	50
6	71	54
7	76	58
8	79	62

TABLE I

Handwriting Records—1917-18—Miss Dora K's Room—Fairview School (Black line means pupil took regular writing instruction; space means pupil exempt)

Group I,	Nov.–1917		Dec.	Jan.	Feb.	Mar.	April	May 1918	
Pupils	Sp	Qual						Sp	Qual
C. L.	80	72.5						86	82.5
H. E.	72	70						78	60
J. R.	76	55						77	60
L. M.	65	60						70	62.5
D. M.	63	67.5						86	57.5
D. S.	78	52.5						88	62.5
G. W.	81	50						78	52.5
H. A.	72	50						75	60
M. T.	75	50						69	47.5
C. V.	91	50						81	62.5
M. B.	60	57.5						77	60
R. W.	62	55						75	62.5
T. B.	53	50						63	45
E. F.	53	57.5						57	52.5
M. H.	54	60						88	62.5
C. M.	52	50						78	60
M. G.	47	55						74	65

C. H.	41	62.5	67	60
L. S.	52	52.5	91	52.5
E. H.	60	52.5	88	62.5
C. N.	53	57.5	76	62.5
M. V.	72	45	72	65
J. R.	76	40	57	52.5
F. M.	74	35	76	42.5
H. D.	63	47.5	70	60
W. S.	65	42.5	88	45
P. G.	67	40	77	50
G. C.	65	40	70	50
A. S.	63	37.5	70	45
E. K.	53	40	73	55
A. D.	57	37.5	74	50
R. W.	46	45	73	60
R. K.	54	37.5	54	37.5
H. C. K.	57	42.5	75	40
R. B.	60	40	66	42.5
J. P.	60	35	69	45
R. H.	66	40	62	52.5
M. N.	63	40	105	52.5
G. K.	56	35	87	47.5
S. W.	54	40	61	52.5
H. M.	56	45	76	47.5
G. M.	52	42.5		
J. G.	74	37.5		

Tables II, III and IV assemble in frequency table form the initial and final scores of the three classes in both speed, or letters written per minute, and quality as measured by the Ayres Scale. Below each table for purposes of comparison are recorded the common measures of central tendency and of variability. Table V summarizes these measures and records the changes effected from November to May in each of the three groups.

TABLE II

Group I. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute-Speed	Number of Pupils Nov. 1917:	Number of Pupils May 1918	Quality of Writing.	Number of Pupils Nov. 1917:	Number of Pupils May 1918
			25–29		
30–34			30–34		
35–39	1		35–39	7	
40–44	1		40–44	10	3
45–49	2		45–49	3	7
50–54	8		50–54	9	11
55–59	4	2	55–59	6	2
60–64	9	3	60–64	3	15
65–69	5	4	65–69	1	2
70–74	4	9	70–74	2	
75–79	4	13	75–79		
80–84	2	1	80–84		1
85–89		7	85–89		
90–94	1	1	90–94		
95–99					
100–104					
105–109		1			
Number	41	41	Number	41	41
Guessed Average	62.5	77.5	Guessed Average	52.5	52.5
True Average	62.95	76.3	True Average	49.94	56.28
Median	62.5	75.96	Median	50.28	54.80
Average Dev.	9.0	7.1	Average Dev.	8.2	6.95
Quartile Dev.	8.5	4.6	Quartile Dev.	7.4	6.24

Growth	Speed	Quality
Absolute	13.46	4.52
Per Cent	21.5	9.0

From a careful study of Table V we may tentatively answer our two questions. The use of the scale did not produce as decided an advantage in Groups I and II over Group III as one might expect, though the increase in speed is enough greater in these

groups to more than balance larger growth in quality in Group III. And it is noteworthy that in every instance the variability of the Groups I and II is reduced more than that of Group III. Perhaps a fairer showing for the use of the scale may be obtained in another way. The fact that Group III started at an abnormally low level of quality, obscures the comparative effect, since the growth was so much easier at that plane. Let us now compare the groups as to increase in the number of pupils up to Ayres fifth grade standards for both speed and quality in November and May. In Group I the number showing fifth grade ability in both speed and quality increased from 9 in November to 27 in May, a net growth of 18 in a class of 41; for Group II there was an increase from 6 to 32, or a net growth of 26 in a class of 44, while in Group III the increase was from 2 to 10, or 8 out of a class of 37. As this is the big task

TABLE III

GROUP II. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute—Speed	Number of Pupils		Quality of Writing	Number of Pupils	
	Nov. 1917	May 1918		Nov. 1917	May 1918
			25–29		1
30–34			30–34	1	
35–39			35–39	1	
40–44			40–44	9	
45–49	2		45–49	12	9
50–54	9		50–54	9	15
55–59	1	1	55–59	9	7
60–64	7	2	60–64	1	8
65–69	7	9	65–69	1	3
70–74		5	70–74	1	1
75–79	5	6	75–79		
80–84	9	10	80–84		
85–89	2	5	85–89		
90–94	1	3			
95–99					
100–104	1	3			
Number	44	44	Number	44	44
Guessed Average	67.5	77.5	Guessed Average	47.5	52.5
True Average	68.98	78.75	True Average	50.34	55.00
Median	67.14	79.17	Median	49.57	54.00
Average Dev.	11.48	8.75	Average Dev.	6.00	5.68
Quartile Dev.	13.06	7.78	Quartile Dev.	5.28	5.15
			Growth	Speed	Quality
			Absolute	12.03	4.43
			Per Cent	18.0	8.9

TABLE IV

Group III. Fifth Grade—Speed and Quality Scores—Initial and Final Tests

Score in Letters Per Minute—Speed	Number of Pupils		Quality of Writing	Number of Pupils	
	Nov. 1917;	May 1918		Nov. 1917;	May 1918
			25–29	3	
30–34			30–34	6	
35–39			35–39	13	6
40–44			40–44	7	11
45–49			45–49	6	10
50–54	2		50–54	1	6
55–59			55–59	1	4
60–64	4	1	60–64		
65–69	2	1	65–69		
70–74	7	6	70–74		
75–79	5	3			
80–84	8	3			
85–89		6			
90–94	3	7			
95–99		4			
100–104	2	2			
105–109	3				
110–114		3			
115–119	1	1			
Number	37	37	Number	37	37
Guessed Average	77.5	87.5	Guessed Average	37.5	47.5
True Average	80.2	88.31	True Average	39.39	46.28
Median	78.5	88.75	Median	38.65	45.75
Average Dev.	11.62	10.54	Average Dev.	5.13	5.0
Quartile Dev.	6.97	9.43	Quartile Dev.	4.51	4.58
			Growth	Speed	Quality
			Absolute	10.25	7.1
			Per Cent	13.1	18.4

TABLE V

Changes Effected From November to May

	Speed		Quality		Variability in Speed		Variability in Quality	
	Average	Median	Average	Median	A. D.	Q.	A. D.	Q.
Group I	62.95 to 76.3 +13.35	62.5 to 75.96 +13.46	49.94 to 56.28 +6.34	50.28 to 54.80 +4.52	9.0 to 7.1 −1.9	8.5 to 4.6 −3.9	8.2 to 6.95 −1.25	7.4 to 6.24 −1.16
Group II	68.98 to 78.75 +9.77	67.14 to 79.17 +12.03	50.34 to 55.0 +4.66	49.57 to 54.0 +4.43	11.48 to 8.75 −2.73	13.06 to 7.78 −5.28	6.0 to 5.68 −0.32	5.28 to 5.15 −0.13
Group III	80.2 to 88.31 +8.11	78.5 to 88.75 +10.25	39.39 to 46.28 +6.89	38.65 to 45.75 +7.10	11.62 to 10.54 −1.08	6.97 to 9.43 +2.46	5.13 to 5.0 −0.13	4.51 to 4.58 +0.07

of fifth grade writing we may answer our first question by stating that the use of the Ayres Scale did add materially to the bringing of a large percentage of the class up to fifth grade standards, and in addition, decreased the variability of Groups I and II.

As to the effect of excusing from formal instruction proficient pupils showing sixth grade ability, comparison of Groups I and II shows that far from retarding grade progress, it increased it in both speed and quality. If in addition to this favorable showing we examine the economy of time for these proficient pupils, the case seems clearer.

Number of Pupils Excused Each Month in Group I, and the Average.

Month	Nov.	Dec.	Jan.	Mar.	Apr.	May	Aver.
Number	3	9	2	7	9	17	6.8

As this saved for each pupil one hour per week, four hours per school month, or twenty-eight hours for the seven months, the time economy was 6.8x28 or 190.4 pupil hours. May we then say that the evidence from these fifth grade tests is in favor of making use of handwriting scales in the class room and of exempting proficient pupils from needless writing practice, thus effecting, with no loss, a saving of time for other work?

Results with Sixth Grade Classes

The work of the two sixth grades is somewhat surprising in that we find them both losing in quality of writing after a year's instruction and practice. The writer feels, however, that this is due somewhat to an abnormal effort in the final May tests to "speed up," with a consequent decrease in quality, as is shown by the median results. Again, it was felt that the elements throughout the year with these groups were not as carefully controlled as with the fifth grade group. The frequency tables for the two sixth grade groups are omitted, but Table VI summarizes the measures of central tendency and variability in which the changes effected from November to May are shown.

The results in Table VI allow us mainly a choice between two evils, which did the poorer? What "works" with one class does not of necessity work with another. The interest, sympathy and enthusiasm of teachers differ and this is apt to be reflected in the groups of pupils. It is of interest however to call attention to two facts, one, that the two classes were somewhat widely different in quality of writing, Group II being of about third grade ability, on the average, and Group I of seventh or eighth grade ability, and at the close of the year, poor as their showing was, this latter group

was up to or beyond the average of sixth graders in general. And this was true after the second fact to be noticed, that is, that 9 pupils on an average were exempt each month from 45 minutes weekly writing, their time being open for other study.

TABLE VI

Changes Effected From November to May

	Speed		Quality		Variability in Speed		Variability in Quality	
	Average	Median	Average	Median	A. D.	Q.	A. D.	Q.
Group I	72.9 to 91.3 +18.4	76.5 to 94.2 +17.7	66.5 to 56.1 -10.4	64.8 to 55.2 -9.6	14.0 to 14.0 0.0	11.5 to 11.1 -.4	8.0 to 9.0 +1.0	6.2 to 9.1 +2.9
Group II	72.3 to 81.9 +9.6	71.4 to 81.4 +10.0	43.9 to 39.4 -4.5	42.0 to 38.3 -3.7	10.3 to 15.4 +5.1	8.5 to 16.9 +8.4	12.4 to 8.3 -4.1	10.5 to 7.1 -3.4

Number of Pupils Exempt Each Month in Group I, and the Average.

Month	Nov.	Dec.	Jan.	Feb.	Mar.	Apr.	May	Av.
Number	14	4	3	7	13	11	11	9

Further than this 19 pupils in Group I were up to sixth grade standards in November and this increased to 22 in May. In Group II, 2 pupils were up to sixth grade standards in November and this decreased to none in May.

While it is recalled that the results with the three fifth grades, where the conditions were somewhat more carefully controlled, were more positive, still it is far from the purpose of those conducting the tests to "prove" anything. The question is simply raised: Of what good is it to require pupils with the speed and quality in writing of grades ahead to take formal practice in writing with the remainder of the class, if they maintain such, and if the time might be employed with other studies in which they may be deficient? We may all have our opinions. These, however, may be set up in a "balance of powers." Actual data obtained under controlled conditions will allow us to arrive at a more positive, if not a more peaceful, conclusion.

COMMUNICATIONS AND DISCUSSIONS

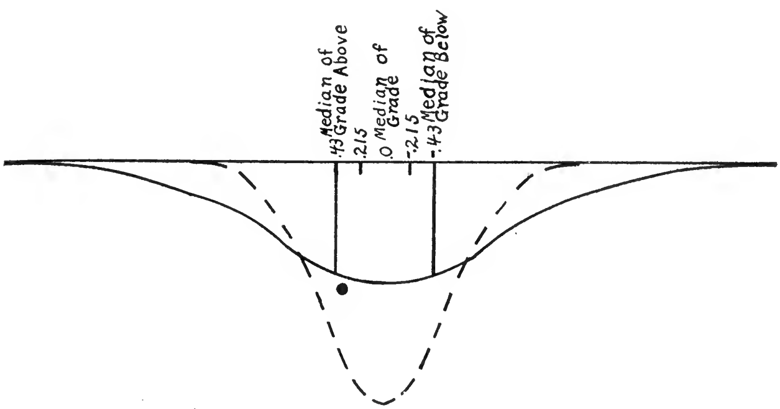
THE MEASUREMENT OF OVERLAPPING

All who have given standardized tests or have used scales for the measurement of accomplishment of children in different grades have found great variability in each grade and extensive overlapping of grade upon grade. The existence of extreme overlapping in reading ability is reported by Daniel Starch[1] who writes "it is found that in speed and comprehension 31.8 per cent of the pupils of any grade reach or exceed the median of the next grade above," and by F. J. Kelly, who reports[2] that, 'The reading ability possessed by the median child in any one grade is superior to the ability possessed by at least one third of the children in the class above him and is no better than the reading ability possessed by at least a third in the class below him."

Assuming a normal distribution of scores in reading ability and calling the standard deviation of such scores Σ, this statement would imply, as may be determined by reference to a table of the frequencies in a normal distribution, that one grade median is $.43\Sigma$ distant from the next. The situation is pictured in the accompaning full-line graph.

[1]Daniel Starch. *Educational Measurements,* p. 42.

[2]*The Kansas Silent Reading Tests,* J. Ed. Psych. v. 7 no. 2, 1916.



Let us assume that the score in the reading test is a valid criterion of ability, and determine the per cent of pupils who would be better classified if they were in the grade above or the grade below. Any pupil making a score greater that that which corresponds to .215Σ is closer to the median accomplishment of the grade above than to that of the grade in which he is located, and therefore should be transferred to the grade above. The percentage of cases in a normal distribution lying above .215Σ is 41.5, and similarly 41.5 lie below—.215Σ. Accordingly, by this criterion, 17 per cent are properly and 33 per cent improperly classified. As, thruout a number of grades, reading ability is presumably the most important single determinant in classification, so extreme a finding as this may well be questioned and carefully considered before finally accepted.

We may approach this matter from another point of view. If we give a five minute test in solving puzzles we are testing ability to react to novel situations and may call our test a reasoning ability test. From data at hand the writer knows that enormous overlapping would be found with such a test. The reason is here quite obvious. The puzzle ability of an individual may be quite identical with his reasoning ability, but it cannot be determined in five minutes. The probable error of a five minute determination of puzzle ability age is very likely three or four years. The great overlapping in this case is seen to be a consequence of the unreliability of the test, and not the trivial nature of the function tested or of the real overlapping of talent from grade to grade.

To a degree the same situation maintains in the case of every measure of overlapping, unless it be in dealing with such measures as height or weight which can be "exactly" measured, and the unreliability of measures must be known before significant estimates of overlapping can be made.

The writer has found in the case of two classes of about thirty, that the reliability coefficient (the extent to which the test correlates with a second similar test) of both the Starch Reading test and the Kansas Silent Reading test was between .2 and .3. More extensive determinations of these reliability coefficients would probably not change the values appreciably, so it is clear how it has happened that such great overlapping was found.

Knowing the reliability coefficient, it is possible to determine what the overlapping would be if measured by means of a very large number of such tests. We will call this the true overlapping with respect to the ability tested. If the scores of the members of a class are the average scores based upon a very large number of tests, all similar to each other in difficulty and in what each tests, the scores will truly represent ability and whatever overlapping is then found will be overlapping in true ability. Consider the dotted curve shown to represent the distribution of scores based upon averages of a large number of tests. If the grades received in the first test are designated by x^2 (x^3 in the second test, x^1 in the third,

x_n in the n'th), where x^1 is a deviation from the group mean, then the full line curve represents the distribution of grades x^1 received by the member of the class (consider N to be the number of individuals), and the dotted curve represents the distribution of the N average grades, each of which is of this type, $\frac{x_1 + x_2 \cdots + x_1}{n}$. We wish to know the spread or dispersion of these latter measures and will therefore calculate their standard deviation:

$$\sigma^2 \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{\Sigma (x_1 + x_2 + \cdots x_1)^2 n}{n}$$ in which Σ indicated a summationof N terms.

$$\frac{\Sigma x \Sigma_1^2 + \Sigma x_2^7 + \cdots \Sigma x_{17} + 2\Sigma x_1 x_1 + 2\Sigma_1 x_2 + \cdots 2\Sigma x_1 x_2 + 2\Sigma x_1 x_2 + \cdots}{n_7 N}$$

Since according to hypothesis the tests which are averaged are all similar they all have the same standard deviation and correlate with each other to the same amount; that is to say

$\Sigma \sigma x_{12} = \Sigma x_2^7 = \cdots = N \sigma_1^2$, according to the usual definition of standard deviation, and

$\Sigma_1 x_2 = \Sigma x_1 x_3 = \cdots = N r \sigma_1 \sigma_2$, according to the usual definition of a product-moment coefficient of correlation.

In this last equation r is the correlation between one text and a second similar one. In the equation given there are n terms such as Σx_2^1, and $\frac{n(n-1)}{2}$ terms such as $\Sigma 2 x_1 x_2$, so that we have

$$\sigma^2 \frac{x_1 + x_2 + \cdots x_n}{n} = \frac{n N \sigma_1^2 + n(n-1) N r \sigma,^2}{n^2 N} = \sigma_1^2 \left(\frac{1}{n} + r - \frac{1}{n} r\right)$$

As n becomes indefinitely large the expression in paranthesis becomes r, so that representing the standard deviation based upon a very large number of tests, *i. e.* the true standard deviation, by σt we have:

$\sigma t = \sigma_1 \sqrt{r}$ (Formula giving the standard deviation of the average of an indefinitely large number of similar tests, knowing the standard deviation and coefficient of reliability of the single tests)

In the case of the Kansas Silent Reading tests, taking the reliability coefficient = .25 and the standard deviation σ_1, we have

$$\sigma t = \sigma_1 \sqrt{.25}, \text{ or } \sigma t = .5 \sigma_1$$

In the original distribution of scores it was necessary to go out from the mean a distance of .215σ to reach a score half way between the median scores of two grades. It will accordingly be necessary in

terms of the distribution of true scores to go out $\frac{.215}{.5}$(=.43) standard deviations to reach the point midway between grades and to go out .86 0 to reach a neighboring grade median. Therefore according to the overlapping reported by F. J. Kelly there is 33 per cent, instead of 17 per cent, of correct classification and, referring to the frequencies given in a normal distribution, 19.5, instead of 33 per cent, exceed or fall short of the median accomplishment of the grade above or below.

All measures of overlapping thus far reported are incorrect as indicative of the numerical amount of overlapping in true ability. The correction in the measure of overlapping here derived is very easy to apply; a table of the frequencies in a normal distribution and the coefficient of reliability being the only new data required.

It is seen that great overlapping is as much an indictment of the test as it is the adequacy of the classification, unless it has been determined that the test itself is highly reliable. Three factors make for overlapping: (a) unreliability of tests; (b) unimportance of function tested; and (c) failure to properly classify, and only in so far as the first two are known not to be the causes can improper classification be established.

TRUMAN L. KELLEY.

Teachers College, Columbia University.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 pei year of nine monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X APRIL, 1919 NUMBER 4

EDITORIAL

THE CLASSICS

Perhaps the war has not changed the fundamental aspects of human thinking as much as optimistic or pessimistic writers of fiction would have us believe, but it has certainly encouraged many people to take account of stock and has made them less tolerant of and complacent with traditional practices. There is a tendency to examine all values anew, and to inquire what each course of action contributes to the sum of human welfare: The failure of the Students' Army Training Corps movement (for it seems pretty generally acknowledged to have been a failure) may have been due to the impossible and unwarranted demands made upon the colleges, but it has afforded the critics of the old time methods and course of study, whether within or without the college, opportunity for renewed and effective attacks upon the uselessness and inanity of much traditional college work. These the colleges, made doubly sensitive by the increased cost of living and by depleted incomes, have keenly felt, and there are many indications of a distinct effort to meet popular demands. Even before the war there was a growing skepticism as to the educational value of

requiring Latin and Greek of all candidates for the bachelor of arts degree, and in most institutions Greek had already slipped into the position of an elective. Since the war this movement seems to have been accelerated, and the announcements of some of our most conservative colleges this spring show a noteworthy reduction in the requirements of the classics and mathematics. For example, Princeton has definitely abolished the requirement of Greek for either entrance or graduation, and has reduced the requirement of Latin to one year during the college course. Yale has made it possible to enter the university without any Latin at all, and has removed many of the advanced mathematics requirements for entrance to the Sheffield Scientific School. This moves an editorial writer in the *New York Times* to bewail the "notable advance of the movement for dehumanizing the humanities and deliberalizing the liberal arts," and leads the noted English classical scholar, Gilbert Murray, into an error of exaggeration. Writing to the *London Times*, he says: "A few months ago a strange and desolating portent occurred. All the American universities at one blow abolished Latin as a necessary subject for the arts degree. Presumably Latin will soon be as moribund as Greek in America."

While one is moved to smile at the idea of American universities being sufficiently united to do anything "at one blow," it is evident that the defenders of the vested rights of the classics to a monopoly of liberal education have reason to be alarmed. This alarm has resulted in the organization of the American Classical League and a plan for an aggressive campaign for the defense of classical learning. Let us hope that some of these champions of the classics may be induced by their ardor for the cause to make statistical and experimental studies on the processes involved in learning Latin and Greek, that the store of our knowledge in educational psychology may be enriched. Perhaps they may succeed in showing us what there is in the technical study of Latin vocabulary and sentence structure that is so remarkably liberalizing and humanizing. In the meantime let us press on in our efforts to work out in secondary schools and colleges a series of broadly liberal courses of study that will offer to each pupil the opportunity to come into contact with the biological, social and political history of the race (including the contributions of the Greeks and the Romans), and will give him training in applying this knowledge to contemporary problems.

J. C. B.

NOTES AND NEWS

Commenting on Mr. Maurice A. Garfinkel's study of *The Effect of the Summer Vacation on Ability in the Fundamentals of Arithmetic*, in the January issue of THIS JOURNAL, Dr. Arthur W. Kallom, of the department of educational investigation and measurement, Boston, writes that he has found among Boston pupils a similar deterioration in scores during the summer months. The differences between spring and fall scores are not so large, but a possible reason for this is the fact that the tests were given in April and October instead of in June and September. Tests given immediately before and immediately after the vacation would bring out more definitely any change due to that period than tests given several weeks before and after. Dr. Kallom deplores the use of the term "median variation" in Mr. Garfinkel's article for the difference between the June median score and the September median score. The term "median variation" has a definite significance in educational statistics, indicating the median of the variations of a group of scores from the central tendency of the group, and it only creates confusion when such a term is used in an obviously different and incorrect sense. In a new science, as that of educational measurements, it behooves us to be particularly careful in the use of our terminology in order to avoid ambiguity. So keenly is this need felt that the National Association of the Directors of Educational Research now has a committee at work on the standardization of statistical terms.

Dr. William H. Burnham, President of the Massachusetts Society for Mental Hygiene, has issued an interesting pamphlet stressing the need for an intensive study of the physical and mental characteristics of each child at the beginning of its school career. He suggests that the first year or two of school life should be devoted chiefly to tests of physiologic age; mental age; habits of observation, attention, association, and the like; general health conditions; and the development of the child's physical resources through play and other forms of motor activity. He hazards the view that there are two marked types of learners, one that learns easily by instruction, by hearing and seeing what others do, by definite directions and demonstration, that is, by the usual didactic methods of the school; and the other a type that learns chiefly by doing, and this by the child's own method. For the latter type the ordinary school procedure is deadly. It is of the utmost importance that these two types be distinguished as early in school life as possible, and the beginning of diagnosis should be at the entrance to school. While Dr. Burnham's assumption of sharply defined types does not seem to be in accord with the results of most experimental investigations, which show continuous variation in every trait studied, this in no way weakens the force of his appeal for a more intensive study of school entrants.

Dr. L. W. Sackett, of the University of Texas, Austin, Texas, has recently issued in neat pamphlet form the Bell-McCollum tests in American history. The eight-page pamphlet includes not only the tests themselves, but the norms derived from their original application, and a weighted evaluation of each question on the basis of those results. The tests may be obtained from Dr. Sackett at the rate of $1.50 per hundred, postage extra.

The annual meeting of the National Association of Directors of Educational Research was held at Chicago, February 27. The forenoon session was devoted exclusively to reports by the members on various technical researches in measuring intelligence and in reading, spelling, geography and composition. The afternoon program covered a more general field, centering about the need for, and the organization and support of educational research work. Dr. B. R. Buckinghan, director of the Bureau of Educational Research, University of Illinois, was elected president, and Dr. Ernest J. Ashbaugh, director educational extension service, University of Iowa, secretary-treasurer.

The March meeting of the New York Society for the Experimental Study of Education was devoted to the consideration of "Problems in the Education of Exceptional Children." The program consisted of two parts. In Part 1, *The Exceptionally Dull Child*, Miss Elizabeth E. Farrell, inspector of ungraded classes, discussed the "After-care of Ungraded Class Children," and Mr. George H. Chatfield, assistant director of attendance spoke on "Problems in Compulsory Education." In Part 11, *The Exceptionally Bright Child*, Miss Anna Gillingham, Ethical Culture School reported on "Intellectual Brightness and School Progress," and Principal Louis A. Marks, Public School 64, Manhattan, spoke on "Intelligence Tests as a Basis for School Grading." The discussion was led by Dr. John L. Tildsley, associate superintendent in charge of high schools, and Professor James E. Lough, of New York University.

VOLUME X MAY—JUNE, 1919 NUMBER 5-6

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE BRIGHT CHILD AND THE SCHOOL

ANNA GILLINGHAM

Principal, Open Air Department,
Ethical Culture School, New York City.

Few problems more perplexing present themselves in school administration or pedagogical procedure than that of grading pupils. How are we to separate them into groups capable of working together for a year with the maximum of individual and collective efficiency? One needs but to glance at the class list of any school grade to be impressed with the great disparity of age. Even in carefully graded schools where considerable effort is made to eliminate the "exceptional," one may find a three or four year range. Curiously enough the youngest pupil is often found to be doing the best work, the child of nine attaining more satisfactory results than his eleven year old classmates, the little ten year old more than holding his own among twelve and thirteen year old youngsters.

Naturally a large number of instances can be explained by fixed grading systems. The child who for any reason, such as ill health, change of residence, or the like, entered school late would have to spend a year in each grade and so could not make up lost time. Yet above all this we see clearly the difference in mental development not commensurate with bodily age.

A few years ago when we became interested in mental tests we felt that they would greatly help us in attaining justice in grading. Since it seemed obvious that children of one grade should have approximately the same mental age, we could test them and classify them upon that basis,

irrespective of physical age and with far less regard than formerly to facts learned. The child of given mental age would be able to do the work of that age, rapidly absorbing any facts of technique which had not chanced to fall in his way in the past. If it were discovered that in a given class there was great disparity of mental age, this fact would reflect some sort of discredit upon the school grading, show very little adaptation to needs of individuals, etc.

The following figures are summarized after giving intelligence tests to six classes in a large private school where the ratio of the number of teachers to the number of pupils was large, the grading reasonably flexible, and a good deal of individual work provided. In testing the first three grades the Goddard, 1911, revision of the Binet-Simon tests was used. With the last three classes the Stanford revision.

Grade IV.		Grade V.		Grade VI.	
Chron. Age. Children	Ment. Age Children	Chron. Age Children	Ment. Age Children	Chron. Children	Ment. Age Children
3...8 yrs.	5...9 yrs.	11..10 yrs.	3..10 yrs.	1...9 yrs.	9..11 yrs.
9...9 yrs.	9..10 yrs.	3..11 yrs.	7..11 yrs.	5..10 yrs.	7..12 yrs.
7..10 yrs.	7..11 yrs.	1..12 yrs.	5..12 yrs.	8..11 yrs.	3..13 yrs.
3..11 yrs.	2..12 yrs.			4..12 yrs.	
1..12 yrs.				1..13 yrs.	

Grade V.		Grade VI.		Grade VII.	
Chron. Age Children	Ment. Age Children	Chron. Age Children	Ment. Age Children	Chron. Age Children	Ment. Age Children
3....9 yrs.	3..10 yrs.	1...9 yrs.	1...9 yrs.	1..11 yrs.	1..12 yrs.
9...10 yrs.	4..11 yrs.	1..10 yrs.	6..13 yrs.	7..12 yrs.	4..13 yrs.
10..11 yrs.	7..12 yrs.	8..11 yrs.	5..14 yrs.	4..13 yrs.	5..14 yrs.
2...12 yrs.	3..13 yrs.	8..12 yrs.	6..15 yrs.	8..14 yrs.	7..15 yrs.
	5..14 yrs.	4..13 yrs.	2..16 yrs.	1..15 yrs.	4..16 yrs.
	2..15 yrs.		2..17 yrs.		

"Poor' Poor"! says the intelligence test enthusiast. "Think of the ineffective work being done. Consider more especially the time wasted by those children who are working with others far below their mental age." It is for them that the most earnest plea is made. Compare for example the following arraignment. The sentences are quoted from Terman because his book happens to be at hand and is especially emphatic, but many others would agree.

"The future welfare of the country hinges in no small degree upon the right education of these superior children."

- - - - - - " Through the leveling influence of the educational lock-step such children at present are often lost in the masses."- - - - - - "The common saying, 'genius will out' is one of those dangerous half-truths with which too many people rest content."- - - - - - "Even genius languishes when kept over-long at tasks that are too easy."- - - - - - "They run the risk of falling into life-long habits of submaximum efficiency. The danger in the case of such children is not over-pressure, but under-pressure."

These statements are all based upon the undebatable assertion that work should be measured out "for each child in proportion to his mental ability," it being assumed that the academic work most needed by each child is that planned in the school curriculum for pupils of the same physical age as the particular pupil has attained in mental age, *e.g.*, a girl of eleven of mental age sixteen ought to be in the second year of High School instead of in the sixth grade where she was found. Very many cases like this are cited in Terman's writings and elsewhere, the ignorance and blindness of the teachers in failing to recognize marked superiority and their unwillingness to allow the child to reap the benefit of it by being pushed ahead being held up for condemnation. The child of superior mind may be advanced for his age, but he is seldom found advanced to the degree warranted by his mental age, etc. etc.

In reply to this arraignment the writer proposes to give a brief description of twenty-five children whose intelligence quotient is 120 or above. To render this array at all significant a few introductory statements are necessary.

The writer who made these tests is a student of psychology and enthusiastic about the assistance afforded by the mental tests. She does not lose sight of the capable child in the routine of grade work if this may be obviated by mental tests, for all the children under her care are given the tests. Although she feels Terman's statements, "It would be desirable to make all promotions on the basis of intellectual ability," (presumably as measured by mental tests) to be too drastic, and that certain knowledge as well as tested power to acquire knowledge is a necessary pre-

requisite to new work, nevertheless she depends considerably upon the help afforded by the tests in making her decisions regarding reclassification. She does not oppose extra promotions for any one of the three reasons set forth by Terman, viz. 1. mere inertia, 2. unwillingness to part with especially satisfactory pupils, 3. traditional belief that precocious children should be "held back" for fear of dire physical or mental consequences. On the contrary she very frequently makes definite provision, even involving radical alterations in program or other routine to accomplish extra promotions. There is unusual opportunity to follow these up and bring them to success because each teacher has every child in the department under her care in her own subject for three years. Much of the formal work is planned on the unlimited assignment basis, so that each child may advance at his own rate untrammeled by the class.

The writer is assisted by three regular teachers, of superior education and ability, who are all in accord with her ideals. In addition there is an unscheduled teacher of special training, whose work is almost wholly with individuals, either in helping the weaker over difficulties or in assisting the stronger to work ahead.

This corps of teachers has charge of a department of sixty odd children of the fifth, sixth, and seventh grades. The school day from nine till three includes two half hour recess periods and one and a half hours of lunch and quiet rest. The program is planned to permit the utmost flexibility, it being possible for a child to recite different lessons in two grades if this seems desirable.

The relations of pupils and teachers in both work and play is unusually intimate, and for the most part there is close sympathy and understanding between home and school so that each can profit by the experience of the other.

Probably little good can be said of any course of study upon ideal psychological grounds. The curriculum of this school is that of other progressive private schools of the country. It is required that the pupils graduated from the

seventh grade of this department shall be ready to handle regular eighth grade work, *i. e.*, any irregularities in the earlier program must be directed to this end.

This department is part of a large school. There is a long waiting list and the standards are high. If it seems probable that a child cannot handle the work of the high school he is eliminated during the elementary grades, except for special considerations of character or unusual gift in some particular line, *e. g.*, music, art, etc. This means that the intelligence quotient has a different relative significance. If the average high school I. Q. is 110 as given by Terman, it will be evident that where the effort is made to receive and retain only those children likely to go through the high school, the average of the grades must be raised considerably above that of the ordinary elementary schools where all are received and natural elimination awaited. Therefore in the following pages it will be necessary to remember that an I. Q. of for example 130 is not relatively as high as when cited by Terman.

The testing was done with extreme conscientiousness in adhering to the standards, and with occasional consultation of more experienced workers. Out of the sixty-seven children of the department two years ago the following twenty-five had an I. Q. of 120 or above. The brief history given covers the school life for three years, the year of the test, and the two following years.

1. Girl. Gr. VI. Age, 11 years. 7 mo. Mental age, 14 yrs. 5 mos. I. Q. 124. Good home. Cultured parents. Undersized and immature in all ways. Freakish and very sentimental, undoubtedly due to approaching puberty. Intends to be a novelist. Sleeps with note book under pillow ready to record inspirations. Gets a notion that a certain subject is hard and fails in it for weeks, then as soon as this improves conceives the idea that she never could do some other subject and goes down accordingly. Writes a very large hand, after much correction and copying for days or weeks suddenly adopts a microscopic hand, then begins writing letters separately, l-i-k-e t-h-i-s. Pities herself a great deal and tries to appear stupid. Every year it has been a great struggle to keep her up to a promotion standard, each year growing a little more difficult as greater maturity was expected of the class and she is apparently on a plateau of development, from which she will in all probability

rise with great suddenness after two or three years. Last year she failed in all thought work, seeming like an immature baby in class discussions. She was taken out of regular work about the middle of the year and given supplementary work. This year, repeating the seventh grade she is doing average work and a great deal of quiet thinking. Delicate, subject to colds.

II. Girl, Gr. V. Age 9 yr. 6 mo. Mental age, 11 yr. 10 mo. I. Q. 125. Very tiny and shy. Like a baby in her point of view. Good mental reactions but very slow. I. Q. would be higher were it not for some of the time tests. Repeated the fifth grade last year because ail subjects were below promotion standard. Used to burst into tears when asked to recite and declare that it was too hard. This was partly due to the fact that the fifth grade year when the test was given was her first year in school, due to delicate health in early childhood. This year in the sixth grade she is doing strong work and is much respected by the class.

III. Girl, Gr. VII. Age, 12 yr. 9 mo. Mental age, 16 yr. I. Q. 125. Father a professional man in responsible executive position. Child beautiful in body and character, loved by everyone, a dreamer and a poet. Almost impossible to keep her down to definite statements or exact conclusions. Mathematics a year long struggle with many failures and "conditions." Grammar difficult, history dependent upon type of work and personal mood. Not lazy, but so dreamy that constant pressure must be exerted if she is ever induced to look beneath the surface. Her eighth grade record was not different from all before. She is now repeating her first half year of high school because of many weeks lost from influenza.

IV. Boy, Gr. VI. Age, 11 yr. 3 mo. Mental age, 14 yr. 2 mo. I. Q. 126. Extreme neurotic. Every muscle tense. Hand work, including writing, almost impossible. Hands almost like those of a paralytic. Has had expert instruction to gain muscular control. Repeated Gr. V. Sees point of problem or situation quickly, but is so verbose and inconsequential that he cannot stick to any one topic but rambles on and on till stopped. No idea of brevity or organization. Covers half a page with a one process problem. Has poor idea of sentence structure, often having several consecutive groups of words with no verb. Naturally it is very hard for him to learn the use of even the commonest punctuation marks. Failed in all seventh grade subjects except geography and was given a special supplementary program. This year is repeating the seventh grade and is doing average work, although still not fully in control of himself.

V. Girl, Gr. VII. Age, 12 yr. 3 mo. Mental age, 15 yr. 4 mo. I. Q. 125. An all round satisfactory pupil in all respects. Her somewhat delicate health and the fact that she takes all work rather

slowly and must feel a sense of thoroughness if she is to be happy in her work and not worry, have made it never seem possible to advance her with any greater speed. Gives much time to music, which is to be her life work. In high school as in eighth grade and earlier elementary she is always charming, always successful, always beloved.

VI. Boy, Gr. VI. Age, 12 yr. 5 mo. Mental age, 15 yr. 6 mo. I. Q. 125. A good natured, faithful lad, plodding away at whatever he is told may help him. Entered this school in the fifth grade when it was discovered that he could not read. Parents had been vaguely troubled but the teacher had told them that he was "such a fine lad." Seemed unable to connect ideas with symbols and failed in all text-book work. Mathematics good. Repeated Gr. V., with expert individual work in reading. In the sixth grade read aloud intelligently, although haltingly. Good substantial work in the seventh grade, although all English subjects were difficult and he was markedly inaccurate, seldom bringing any task to a conclusion without at least one glaring error. In the eighth grade his work is average, but his depth of interest as a student steadily increases.

VII. Boy, Gr. V. Age, 9 yr. 6 mo. Mental age, 12 yr. 2 mo. I. Q. 128. Son of physician in position of great responsibility. Mother unusually intelligent. Undersized and babyish. In fifth grade never volunteered or made any sound unless forced to do so. So slow in execution that one or two problems or a half dozen sentences represented an hour's work. Gave correct, very brief replies, often showing excellent thought in oral lessons under compulsion and was considered markedly intelligent in home activities. Afraid of rough games of children of his age, preferring little children. In the sixth grade did not seem to have advanced at all in development. Repeated the grade this year and although still very quiet, is doing excellent work and showing the qualities of a real student.

VIII. Girl, Gr. V. Age, 10 yr. 11 mo. Mental age, 14 yr. 1 mo. I. Q. 129. Satisfactory in all subjects, she impressed no one as having any great depth of feeling or thought. Bright, vivacious, much petted by older pupils. Left at end of sixth grade to go to another school.

IX. Girl, Gr. VII. Age, 12 yrs. 4 mo. Mental age, 16 yr. 1 mo. I. Q. 130. Keen and quick, inaccurate and superficial. Skipped one or two grades in earlier years and had to repeat grade VII just to grow up to herself. Was too immature to manage the situation, lost papers, forgot assignments and had no conception of the human import of topics discussed in history, industrial geography, civics, etc. On repeating, did fair work and has progressed in eighth grade and high school so far without "conditions." She excels in algebra and grammar but sees little in literature or history.

X. Girl, Gr. VII. Age, 12 yr. 8 mo. Mental age, 16 yr. 5 mo. I. Q. 130. Strong and charming personality. Great promise of capability. Does creditable work the last quarter of each year, but is so slow in making adjustments to new personalities, topics, vocabulary, etc., that the first part of the year is a struggle against failure. Has poor methods of work; slow and blundering in execution; always loses her place, breaks her pencil, etc. Cannot spell and has great difficulty in acquiring correct terms, amounting almost to Mrs. Malaprop at times. Special difficulty with mathematics. Is repeating the first term of high school due to time lost in influenza.

XI. Girl, Gr. V. Age, 10 yr. 8 mo. Mental age, 13 yr. 10 mo. I. Q. 130. High pitched, babyish voice and superficial prattle. Very keen visualization and imagination. Can learn anything but has no depth of appreciation in literature and does poor work in history and geography through lack of comprehension of relations. Was a problem during primary grades on account of her inattention and lack of respect for the rights of others.

XII. Boy, Gr. V. Age, 10 yr. 8 mo. Mental age, 14 yr. 1 mo. I. Q. 132. Extremely nervous and tense but bright and capable and increasingly successful. Very naive and childish, causes great amusement to other children.

XIII. Girl, Gr. VI. Age, 11 yr. 9 mo. Mental age, 15 yr. 10 mo. I. Q. 135. Unusually capable parents. Child of fine, strong personality. For several years cared only for mathematics and would study nothing else. Declared with tears that she could not study history and geography. Had great difficulty in gaining promotion to the seventh grade in these subjects. Months of tutoring, summer work, make-up examinations, etc., sent her into the seventh grade radiantly successful but with her interest in mathematics somewhat alarmingly diminished. Her year in the seventh grade was one of remarkable student activity for a child of twelve, voluntary research work, etc. We greatly regretted it when she entered the high school of another school.

XIV. Girl, Gr. V. Age, 11 yr. 1 mo. Mental age, 15 yr. I. Q. 135. Skipped third grade on general ability but failed flatly in fifth grade because of back work superficially covered and also because her self-satisfaction was so great that she could not be induced to work. Repeated fifth grade well and in the sixth grade did work a little above the average although not as good as several children of a lower grade of mentality. In the seventh grade her complacent indolence causes much trouble. She does poor work till on the brink of failure, then shows her ability by bringing everything up suddenly, only to lapse again. Poor attitude toward classmates.

XV. Boy, Gr. VI. Age, 11 yr. 5 mo. Mental age, 15 yr. 7 mo. I. Q. 136. A great problem at home and at school. Learns easily but is satisfied to recite big words like a parrot and lacks force and depth of character. Written work almost illegible. Very suggestible, he will not stand up for his really good ideas if others demur, and he yields to every temptation. When reproved he looks up with a sweet, vague face to explain how others "did it first." Did not seem to develop at all in three years. Has gone to another school.

XVI. Boy, Gr. V. Age, 10 yr. 4 mo. Mental age, 14 yr. 1 mo. I. Q. 136. Eminently satisfactory in work. Delicate in health. Earnest and painstaking. Left after one year.

XVII. Boy, Gr. VI. Age, 10 yr. 6 mo. Mental age, 14 yr. 5 mo. I. Q. 137. A very clever baby. Undersized and frail in body. Has lacked the grasp on vital relations and human problems which one expects of a boy of his age to say nothing of his mental age. Good in mathematics; literature and history are little but memory work. Has no message for the world in composition. The amusing faults, petty deceits, etc. of a very little boy. Each of these statements has grown less and less true year by year, until now in his first high school year we find him grown considerably, rather less shy, acquiring more friends, and getting a stronger grasp of all work. The same general characteristics still persist, however.

XVIII. Girl, Gr. V. Age, 10 yr. 5 mo. Mental age, 14 yr. 6 mo. I. Q. 139. A sweet frail little girl, very conscientious in all moral situations but lacking pride or conscience regarding the quality of her work. Would not do anything at all better than she was absolutely forced to do. Marks slightly above average in thought, verging on failure in appearance of papers. Slow and inaccurate. There has been a slow, steady improvement, until now her work is seldom criticized, but she is so quiet and unaggressive as to make almost no impression upon class sentiment or opinion.

XIX. Girl, Gr. VII. Age, 11 yr. 10 mo. Mental age, 16 yr. 9 mo. I. Q. 142. A marvel of efficiency in all directions, perfect health, remarkable speed in execution, remarkable maturity of thought and perception of inner relations. Not popular with classmates, who are annoyed and amused by her naive self-esteem and childish exuberance of joy in her own success. Now in high school. Probably her adjustments are better, but there is still painful disparity between her immature impulses and conceits and her maturity of intellectual grasp.

XX. Boy, Gr. V. Age, 10 yr. 3 mo. Mental age, 14 yr. 11 mo. I. Q. 146. Remarkably intelligent parents. Always a drag in class, making promotions with heavy conditions. Withdrawn at end of first year to join very small class. Undersized. Droll, old

way of talking. Remarkable thought when not hurried. Incapable hands so that written work was a failure. Poor speller. Slow in calculation so that oral arithematic was poor. Almost incapable of getting an explanation or following a direction unless given very slowly to him alone. Looked dazed most of the time in class and replied at random.

XXI. Boy, Gr. VI. Age, 11 yr. 6 mo. Mental age, 16 yr. 11 mo. I. Q. 147. Excellent mind for grasping full import of very mature thought but no student quality. Gets a great deal in brief time but never the most because he will not work for it. All this was true in sixth grade. In high school he seems to be attacking the wrok with slightly more pride and strength.

XXI. Boy, Gr. VI. Age, 11 yr. 11 mo. Mental age, 17 yr. 8 mo. I. Q. 148. Parents not of exceptional ability. All round in attainments but not at all above level of grade. Too conceited to put forth great effort. Required a good deal of prodding to keep him up at all. Left at end of first year.

XXIII. Boy, Gr. VI. Age, 11 yr. 9 mo. Mental age, 17 yr. 4 mo. I. Q. 148. Good in all lessons. Conspicuous for his initiative and originality in practical affairs. Faithful in the performance of required tasks, he rather prides himself about not being "roped in" by any scheme of the teacher for getting any extra work out of him. Never volunteers for unnecessary topics. Uses his brains to save himself unnecessary exertion. During the last two years has shown some real pleasure in success in mathematics and willingness to work for that pleasure.

XXIV. Boy, Gr. V. Age, 9 yr. 11 mo. Mental age, 15 yr. 7 mo. I. Q. 157. Moved to another city at end of sixth grade. The most casual observer would have picked him out as a boy of unusual intelligence. His written work was so poor, however, and he made so many false strokes, so often got the wrong answer by the right process, or forgct to read all the problem, or lost his paper, or mislaid his book, or said the wrong date when he knew better, just from general blundering eagerness, to say nothing of the times when exuberant spirits wound up in a time-consuming snarl of some sort, that it was a constant care to his teachers, allbeit amusing and fascinating, to keep him up to the average of his class. Furthermore, his infantile gambols and naive observations rendered ludicrous the idea of class-room companionship with boys of fifteen.

XXV. Gr. V. (VI soon after test was taken) Age, 9 yr. 8 mo. Mental age, 15 yr. 4½ mo. I. Q. 159. A year later I. Q. 170. When this remarkable test was sent to Terman and Goddard two years ago both advised rapid advancement, the latter openly saying that the child should be given work of high school grade.

The boy was practically tutored by the special teacher for half a year, during which time he made up nearly a year's work and was placed in the grade ahead. Here he stayed by dint of relentless prodding on the part of the writer. This he took with the utmost good nature and amusement, even gratitude. "I should never have had the *A* if you hadn't made me." Whenever there was a difficult new subject presented he glowed and sparkled. He got it all ahead of anyone else, but he would not go home and do the patient digging required of the other children and needed by him to fix facts in his mind. He reveled in history and read a great deal, but superficially and failed in daily lessons. He could explain any problem and point out the mistakes of other pupils, but he seldom attained the right result to a problem of any length because he didn't stick at it long enough, ignored some of the conditions, etc. Although he had a marvelous vocabulary and power of imagery and sense of humor his composition was poor because he could not spell and it was so much trouble to write things that he stated them in the briefest and most meagre style. There was always some reason why the task could not be done when assigned till his work dragged behind in a manner to drive an intelligence test idealist to distraction. No one who ever worked with hin for an hour doubted his remarkable ability, but how to achieve results? Tutored he could really become one of the child prodigies, but he surely needs social life, for he is a poor mixer with those of his own age, a poor sport, and his humor and quickness in conversation are suitable for grown up companions rather than for children. Furthermore, he is delicate in health. It was definitely planned to have him take two years for the seventh grade in order to attain a maturity and acquire methods of work which would promise some hope of success in the eighth grade which is a part of the high school department and which cannot make a boy sharpen his pencils and pursue him when he loses his books or needs to wash his face.

In repeating the grade this year under a new teacher he started with failure in all subjects. The teacher could not believe that he had ability. Then she placed him in charge of a project to collect and organize current events clippings. In this distinguished position he has worked most diligently and his general work has improved somewhat.

Here they are, my twenty-five little men and women. What lesson do they teach us? To many the first thought is, "The mental tests are not a reliable index of the child's ability and have small value for us as teachers." "Look at that boy with an I. Q. 125 repeating his grade and this girl with I. Q. 130 struggling to keep up her mathematics." Naturally we must all recognize limitations in quantitative measurements when applied to personality. The

biggest things in life are not measurable. There is no test for character or will power, nor any index which will show that one child is so constructed as to keep much in reserve, while another gives all of himself so that a test reveals the uttermost of which he is capable. Nevertheless the writer is increasingly disposed to attach importance to the tests and to rely upon them for prognostication of a child's future achievements. By and large they tell the truth as to the child's potential ability to grapple with situations.

Nearly all persons using mental tests without regard to physical age have found a good many inconsistencies in their groups. Not only are children with high I. Q.'s found doing poor work, but sometimes those with low I. Q.'s are seen to be doing good work. The former are usually the clever babies who are being compared in steadiness, reliability, and emotional reactions with children two, three, or four years older; the latter are temporarily comparing favorably with much younger children.

Dr. Goddard, Dr. Terman, and others who would have us believe that a fifteen year old boy with an I. Q. of 100, a ten year old boy with an I. Q. of 150 and a twelve year old boy with an I. Q. of 125 can be classed together because their mental age is fifteen years, are ignoring what seems to me the fact that maturity of character, emotions, and instincts do not follow intelligence levels as closely as they do physical development, expecially in the early adolescent period and just before.

It is very desirable that ten year old children with I. Q.'s of 150 should be classed with other ten year old children with similarly high I. Q.'s, not with those of fourteen or fifteen with I. Q.'s around 100. So grouped the superior children can be placed under demands suitable for them and their reactions will invariably be very different from those of low I. Q.'s of the same chronological age.

It seems equally clear, however, that one cannot convert this power with unfailing certitude into academic success along paths uniformly laid out. It is not enough to say with Goddard and Terman that such children are not permitted to do what they might, meaning simply rapidity of advancement, nor that, "Wherever located such children rarely get anything but the highest marks, and the evidence goes to show that most of them could easily be prepared for high school by the age of twelve years."

1. The very excellence of such minds as these renders them individualized and forceful so that many of them will not docilely accept subject matter with no vital bearing for them, depending upon memory and doing the expected thing as the mass of children do.

2. To a very large extent these children are poor in muscular control and hence find writing very difficult. Under the most liberal interpretation our school routine demands a large amount of written expression. Many of them are neurotic and have uncertain control in many directions, little power of inhibition, or strength of will to concentrate.

3. It is characteristic of a large percentage of these children that they see the essential principle at once and have a loathing for detail and elaboration. Hence the elementary years, which are the years for acquiring the tools of education are difficult ones. Such a child says, "I know how to do it, I just made a mistake in the work." In many cases of which the writer has personal knowledge the result has been that scarcely any problem or task was ever carried out to a successful conclusion. Only the first original thinking held the attention; the rest was drudgery to which they would not submit and unless unusual wisdom is employed in the direction of such children, their lives will continue to be freighted with brilliant possibilities and actual failures, since a large part of life must be given to execution. "Allowing" such children to hurry along over failures and get into high school does not offer a solution.

4. There are other sides to life besides the academic. Some children are going to use their high intelligence in meeting practical situations and try as they will they cannot hitch it up to school requirements. As I write I can lift my eyes to a building which is regarded as one of the greatest architectural and mechanical (mathematical) feats of America. The man who drew the plans and worked out the calculations was a boy who could not progress in school, being expecially deficient in mathematics. He educated himself by reading what he pleased and when the need for mathematics came the ability was there. It is not true that the work of such children is always superior in whatever grade they are found.

5. Probably most children of this grade of intelligence could be given the mathematics and language work of the high school several years earlier. But high school courses are planned for adolescents,

and the brilliant pupil of ten or eleven has the instincts and emotions of a little child. He is not prepared to grasp the import of high school literature, history, civics, or ethics or to expand in the adolescent atmosphere, to respond to the student government, the athletic code of honor, the interplay of personalities in the friendships of youth. He has intelligence but he has not yet lived or felt.

With the parenthetical qualification that a large part of the mental and moral poise of a community is due to its average men and women doubtless we should all accept the statement that "the future welfare of the country hinges in no small degree upon the right education of these superior children." "The responsibilities of those who deal with them are very great." The crux of the matter lies in these words "right education."

The almost universal public school verdict all over the country has been, "get the bright children through school as quickly as possible and out at work or into high school or college. Save their time!"

The reasons for this attitude are easy to see. One is economic pressure. The children must get to work as soon as possible. If they are to enter professions they must not delay the beginning of their preparation. Another reason is the nature of the public school requirements. It has been shown that a faithful child can go through the public schools, grade by grade if his I. Q. is 90 or above. If, therefore, one is planning for I. Q.'s of 125 or above with no alteration in curriculum the only thing to be done is to advance them very rapidly. Thus the only advantage accorded by the educational system to the superior mind is the curtailment of his school life!

Is this not offering a stone in the place of bread?

One authority writes; "I have no sympathy, myself, with the cry that our young people must get into higher institutions earlier. Anyone who has had contact with young people leaving home at eighteen years of age to go to college— – – – – will certainly feel satisfied that they are none too old for the adventure." Harvard shows successful academic records for its sixteen year old freshmen, but there are instructors there who feel that they would have received much greater benefit in the larger college life had they prolonged their preparation in more intensive study.

Has the time not come for us to work purposefully on the question of differentiation in course of study for the supernormal children, just as so many people have worked to meet the needs of the subnormal?

There are certain mechanical tasks over which there is no advantage to linger. There is no use in holding back the natural speller or the child who learns to read as easily as to talk. But certainly there should be opportunities given the child of this type who also has superior intelligence to use his ability in more intensive work in history or literature, in more applications of the same principle in mathematics than can be attempted by the child who is struggling with the tools. If the bare requirements of a grade demand the crossing of a field of history by a single path, the gifted child ought to be enabled to wander in that field by many paths; he ought not to be hurried into the next field after the same meagre glimpse. We are always told that it is not the intention to hurry into college those immature in judgment, physical development, instincts, emotions, and the like, that the pupils to be advanced are not merely the keenly intellectual, but those of all round development, mind, body, character. I have sought for such children and I have not found them. I have asked many people in high school and college work, and physicians. The usual answer is, "If they exist, I haven't known them."

Instead of encouraging the bright children to spend all their time on school subjects in order to get ahead, thus becoming still more academicized than they are, might we not use their keener minds in part to reduce the time spent in lessons, supplying instead more of those activities which the intellectual so especially need? First, joyous recreation and vigorous free play, more than they can get in two or three gymnasium periods a week. The physical record of school children is nothing to be proud of, even in the best of schools. Eye defects, neurasthenia, spinal curvatures — — — increase grade by grade. It is true that a majority of the people of distinction and leadership are neurotic. But neurotic children of brilliant promise cannot attain success unless they are helped to sound bodies and as much nerve poise and control as possible. The keen, brilliant child would seem far less in need of intellectual activities such as current events or science clubs than of swimming, long hikes and the like.

Second, more time might be saved in these children's education for music, art, and industrial work of many sorts.

It may reasonably be asserted that the needs of many bright children would be better served by such methods as the above, than by being rushed into college a couple of years earlier with only the same number of technical requirements fulfilled as in the case of the slower minds.

And then what about those other types already cited – – those who have great difficulty with the technical school work of the early years; or those who seem doomed to inevitable failure by their superficiality, and by the very ease with which they can get along somehow? What are their real needs?

MENTALITY AND SCHOOL PROGRESS

HERBERT A. TOOPS AND RUDOLF PINTNER

Ohio State University

A recent very interesting article by Kornhauser (*) makes use of Yule's association coefficient, Q, as a measure of the relationship "between intelligence of children, measured crudely by school advancement, and the economic standing of parents, measured still more crudely." The possession or non-possession of a telephone was used as a measure of the parental standing in regard to wealth. His final association coefficient of +.61 between Advanced and Telephones, Advanced and Not Telephones, Retarded and Telephones, Retarded and Not Telephones was found by disregarding the two classes, Normal and Telephones and Normal and Not Telephones. Had he included all normal cases with the advanced cases, he would have secured a much lower coefficient, Q = +.30 for the four (now different) classes, Retarded and Telephones, Retarded and Not Telephones, Normal and Above and Telephones, Normal and Above and Not Telephones. The size of the coefficient is materially affected by leaving out the normal group, evidently depending upon the number of the normal group eliminated in the calculation of the coefficient. A coefficient so obtained, however, gives us a better answer to the question with which most of us are concerned, namely, "Are the extreme members of the one series also at the same extreme in the other?"

The procedure used in the article cited has suggested to the writers the use of a similar procedure with the data on hand in regard to the relation of school grade to mentality, a perfect or nearly perfect correlation of which was assumed in Kornhauser's investigation. Are there any conditions under which school grade may be taken as a perfectly reliable index of mentality? If there are such, broad problems, such as Kornhauser has dealt with, may be solved without resort to actual mental testing of large numbers of children.

The problem which we set for ourselves then is: Can we in this investigation select our lines of division, by eliminating a middle or normal group either of school grades or of mentality, or both, so that school subnormality shall always mean mental subnormality,

---

*KORNHAUSER, A. W. *The Economic Standing of Parents and the Intelligence of Their Children.* Jour. of Educ. Psych., Vol. IX, No. 3, 1918. pp. 159–163.

and school advancement always mean mental advancement? Or, since we can readily ascertain a child's grade in school, can we find limits such that, to be in a given grade at a given age shall always mean a given mental diagnosis? The conditions necessary for the solution of the problem will have been fulfilled if we can determine limits such that the resulting Q shall be +1.00.

Method and Results

Three schools, A, B and C, were given a series of group mental tests. (*) The children of school A were of superior social status, those of B of medium social status, while those of C were of inferior social status. A median mental age was computed for each child and from this the intelligence quotient, I. Q., calculated. In the absence of directions to the various examiners in that regard, some required the children to give their age on their last birthday, while some gave no directions at all. The best assumption under the circumstances thus seems to be that all children who reported their age as eight were past their eighth birthday. That this is a fair assumption is shown by the fact of our final selection of grades, 3B and 3A, to represent the "normal" grade for eight-year-old children, and by the percentages of school retardation and advancement thereby secured.

The formula for Q, used throughout this investigation, is:

$$Q = \frac{(AB)\ (ab)\ -\ (Ab)\ (aB).\dagger}{(AB)\ (ab)\ +\ (Ab)\ (aB).}$$

The various procedures and the limiting values used in each are given in Table I, Part I, and the resulting Q's are given in Part II of the same table. The conditions of the table are all given as for age 7, the corresponding conditions for all other ages being of course understood. All the procedures include all the ages and the Q in Part II of the Table is for all the ages concerned. Thus, Procedure I includes all ages from 6 to 16. For age 8, the school divisions would be 3B and below, 3A and above; the mental divisions, 7.9 and below, 8.0 and above. For age 9, the school divisions would be 4B and below, 4A and above; the mental divisions, 8.9 and below, 9.0 and above. And so on for all the other ages. The Table merely illustrates by showing the method at age 7, but the Q's are calculated by using all ages.

*Pintner, R. *The Mental Survey.* 1918, pp. 40-50.

†Yule, Udny. *An Introduction to the Theory of Statistics.* 1916, p. 38.

TABLE I.

The Conditions and Limiting Values Used in the Computation of Q by the Various Procedures, and the Q's Derived Therefrom (All conditions given as for the age 7.)

Procedure	I					II
	School		Mentality		Ages	
	A	a	B	b		Q
I	2B and below	2A and above	6.9 and below	7.0 and above	6 to 16 inclusive	+.56
II	1A and below	2B and above	6.9 and below	7.0 and above	6 to 16 inclusive	+.646
III	1A and below	2B and above	6.9 and below	7.0 and above	As of Proc. III	+.654
IV	1A and below	3B and above	6.9 and below	7.0 and above	As of Proc. III	+.85
V	1A and below	3B and above	I. Q. of .89 or below	I. Q. of 1.10 or above.	As of Proc. III	+.97
VI	Retarded 4 grades or more	Advanced 2 grades or more	I. Q. of .89 or below	I. Q. of 1.10 or above	As of Proc. III	+1.00

Procedure I. Massing together all of the 1723 children of the three schools, we first considered all 7 year-old children in grade 2A or above as "Above normal," all in 2B or below as "Below normal." In mentality, all above the median mental age 7.0 were considered mentally "Above normal," and vice versa, and similarly for all the other ages. QI is then +.56, which is a comparatively low value. That this figure does not best represent the relation of the intelligence of school children to their grade in school is shown by the fact that there are more than 73 per cent retarded pupils in the school by this procedure. Hence it was decided that it was probably better to compute Q, using as below normal in school all those children of age 7 who are in grade IA or below.

Procedure II. With the above limit for school "Below normal," we obtain a QII of +.646. With this limit we have 46.2 per cent of below normal children in the three schools. This seems to be more nearly in accord with the percentage we should expect, and furthermore, is more in accord with the common sense view of a 7 year-old retarded child being one who has not yet been promoted to the second grade.

Procedure III. An attempt at computation at this point of the Q values for each age separately discloses the fact that such values cannot be calculated at all for those ages in which all of the pupils are either retarded or advanced by the above criterion. In schools A and B, both of which had junior high school courses or a ninth grade, all pupils of age fifteen or above are evidently retarded by the above criterion; similarly with all those of age fourteen and above in School C which had no ninth grade; and all children of age six are advanced in all three schools. From our total figures, then, should be subtracted all those children who have not an opportunity to be either retarded or advanced in school. If we do not do this, we are evidently not getting the value of Q we should by reason of the fact that at the one extreme the brighter and more advanced children of age fifteen and over are now in high school or at least graduated from the grade school, while at the other extreme, we have too many advanced pupils by reason of the fact that a six-year-old cannot be retarded at all because of the ordinary practice, assumed in this problem, that the age at entering school is six years.

Omitting, then, those ages in which all children have not a chance to be both retarded and advanced in school *at least one year*—that is, ages, 6, 14, 15 and above in Schools A and B, and ages, 6, 13, 14 and above in School C—gives us a total QIII of +.654. From Table II, which gives the Q—values by ages, it is quite apparent that the adjustment of mentality to school grade by means of re-

TABLE II.

Q—Values by Ages and By Schools for Procedures, III, IV and V.

	Calculated Q—Values:—										
	Age:—							School:—			
Procedure	7	8	9	10	11	12	13	A	B	C	Total
III	—.25	+.73	+.91	+.90	+.60	+.81	+.66	+.49	+.67	+.83	+.654
IV	+.34	.90	.96	1.00	.70	.96	.76	.78	.79	.94	.85
V	+.29	1.00	1.00	1.00	.87	1.00	1.00	.91	1.00	.996	.97

tardation has barely begun at age 7. By omitting age 7 also we secure QIII a = +.72. This, however, seems unfair as this age, which can be retarded in school not more than one year, should be left in to compensate for the age at the other end of the series which cannot be advanced in school more than one year. Accordingly

age 7 is included in all further computations, while all ages which have not the possibility of being retarded and advanced in school at least one year have been eliminated as outlined above.

Procedure IV. An inspection of a distribution table of children by years of retardation and advancement showed conclusively that the Q—values are not higher than they are chiefly because of the variation of the group of "normal" school children. We, therefore, proceeded to calculate Q with the school "Normal" children eliminated— that is, with all 7-year-old children in grades 2B and 2A eliminated, and correspondingly for the other ages. We thus secure QIV = +.85. This is to be regarded as a coefficient representing the relation for all three schools combined of the four classes: School Retarded and Below the Median in Mentality, School Retarded and Above the Median in Mentality, School Advanced and Below the Median in Mentality, School Advanced and Above the Median in Mentality.

Procedure V. While the above coefficient, QIV, is sufficiently high to indicate a very high degree of relationship between mentality and school grade, the coefficient is not such as to warrant our using school grade as a prediction of mentality. If, by eliminating from our mentality distribution also a middle class such that we secure a nearly perfect Q for the four classes, School Retarded and Mentally Retarded, School Retarded and Mentally Advanced, School Advanced and Mentally Retarded, School Advanced and Mentally Advanced, then we may say that such-and-such an amount of school retardation almost certainly means such-and-such a degree of mental retardation as determined by the limits finally assigned. To secure comparable measurements of mental retardation at different ages, the I. Q. values were computed. Various limiting I. Q. values for the middle class were tried out. The final selection coincides, in all but the division between the bright-very bright, with the C. I. A—values used by the writers with the Point Scale for making diagnoses into the groups, "Feebleminded, Backward, Normal, Bright, Very Bright." The limits used and the percentages falling in each diagnostic group are shown in Table III. The resultant percentages show this selection to be reasonably accurate as to the diagnostic limits involved, according to the generally accepted views as to the distribution of I. Q.'s of school children where large groups are taken. This procedure results in a QV—value, when "normal" school and "normal" mentality groups are eliminated, of +.97. The actual distributions in numbers by grades and by diagnosis are shown inTable IV.

TABLE III.

I. Q. Diagnostic Limits, and Distribution of Diagnostes by Percentages.

Diagnosis	I. Q. Limits	percentage:— Procedure III	All Ages
Feebleminded	Up to .69 (inclusive)	1.5	2.0
Backward	.70 to .89 (inclusive)	18.6	21.9
Normal	.90 to 1.09 (inclusive)	51.2	50.3
Bright	1.10 to 1.29 (inclusive)	24.1	22.4
Very Bright	1.30 and above (inclusive)	4.6	3.4
	Total	100.0	100.0

TABLE IV.

Distribution in Numbers of Children, by Diagnosis and by Grades of School Retardation or Advancement. (All schools and ages of Procedure III. A minus sign indicates retardation, and plus indicates advancement. Two grades are included in the "normal" group.)

Grade	Number at Diagnosis:—					Total	
	F. M.	B.	N.	Br.	V. Br.	No.	Per Cent
—8	1	..	..	..	..	1	0.1
—7	2	..	..	..	..	2	0.2
—6	3	4	..	..	..	7	0.6
—5	5	8	5	..	..	18	1.5
—4	1	24	7	..	..	32	2.6
—3	1	36	25	4	.	66	5.4
—2	2	49	74	13	..	138	11.3
—1	3	50	94	36	1	184	15.1
N.	..	50	360	147	28	585	48.0
+1	..	6	42	81	18	147	12.1
+2	..	..	12	11	4	27	2.2
+3	..	..	3	2	5	10	0.8
+4	..	..	1	..	..	1	0.1
Retarded	18	171	205	53	1	448	36.8
Normal	..	50	360	147	28	585	48.0
Advanced	..	6	58	94	27	185	15.2
Total:							
Number	18	227	623	294	56	1218	..
Per Cent	1.5	18.6	51.2	24.1	4.6	..	100.0

The formula for Q is imperfect for the reason that, in this case we have a relatively very large "School Retarded-Bright and Over"-group in comparison with the "School Advanced-Backward and Under"-group, and yet Q is nearly perfect. For this reason we have reduced Table IV to a form (Table V) showing by *years* of retardation and advancement just how many chances out of one hundred a child of a given degree of retardation or advancement.

TABLE V.
Chances Out of One Hundred of Falling in a Given Diagnosis, By Years of Retardation and Advancement.

School Years	Chances in 100 of Diagnosis					No. Cases	Per Cent
	F. M.	B.	N.	Br.	V. Br.		
—4	100	..	..	..	..	3	0.3
—3	32	48	20	..	..	25	2.1
—2	2	61	33	4	..	98	8.0
—1	2	31	52	15	..	322	26.4
N.	..	9	62	24	5	585	48.0
+1	..	3	31	51	15	174	14.3
+2	..	..	36	18	46	11	0.9
No. of						Total	100.0
Cases	18	227	623	294	56	1218	..

collectively, has of falling into each diagnosis. By reference to Table V, we see that a 11 year-old child in the third grade (3 years retarded) has about one chance out of three, $\frac{32}{100}$, of being feeble-minded, and four chances out of five, $\frac{80}{100}$, of not being above backward, with absolute certainty, $\frac{100}{100}$, that he is not above normal in mentality.

Procedure VI. By reference to Table IV, we see that no individual who is four or more grades retarded in school is above normal in mentality; also, that no individual who is more than one grade advanced in school is below normal in mentality. With these, limits, than, for our "normal" mentality group, we secure a perfect positive association, QVI = +1.00.

By calculation from the summation figures of Table IV, we find that there are 36.8 per cent. below normal in school, with only 20.1 per cent. below normal in mentality. On the other hand, there are only 15.2 per cent. above normal in school with 28.7 per cent. above normal in mentality. This seems to suggest that there are from 12 to 16 per cent. of the children of the three schools retarded in school and yet possessing mentality such as to enable them to do at least normal work. There is an equal percentage who are only normal in school and whose mentality is such as to enable them to do advanced work. We find 32.9 per cent. of the "normal" mentality group retarded in school, 18.0 per cent. of the "bright" group and less than 2 per cent. of the "very bright" group. Our results,

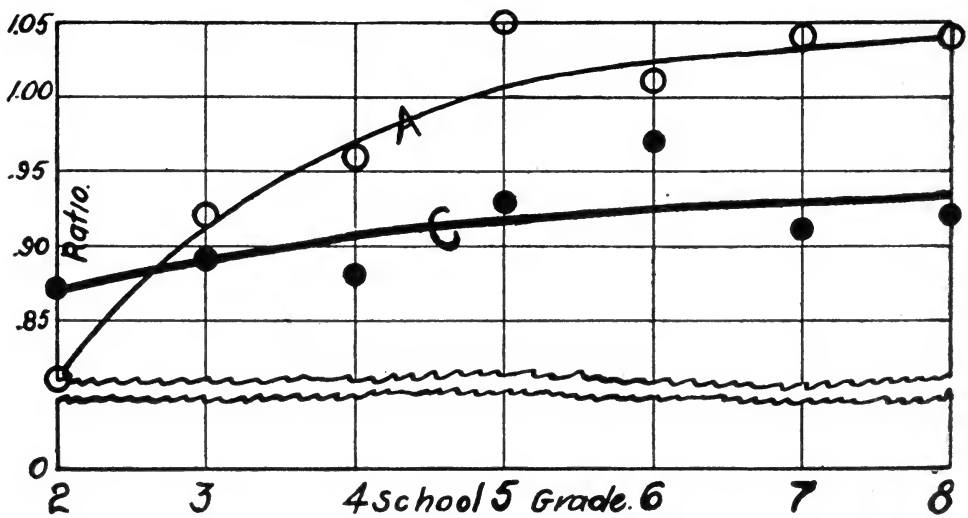
when school performance is measured in terms of school retardation and advancement, may possibly be best interpreted in the light of observations made by Whipple:*

"When ability is strikingly superior, school performance is superior, whether the student is industrious or not. When ability is strikingly inferior, school performance is quite inferior whether the student is industrious or not. When ability is of average amount, school performance is distinctly affected by industry and zeal. Most of the divergences between test results and the school records or the teacher's estimates pertain to the position of these students of average native ability."

May not even the presence or absence of industry and zeal be almost wholly dependent upon the school's willingness or unwillingness to recognize such virtues by the reward of school advancement?

In an effort to secure some light on this question, we have plotted, (Graph I), for Schools A and C, the curves of the ratio of the median mental age at each grade divided by the corresponding median

*WHIPPLE, G. M. *The Use of Mental Tests in Vocational Guidance.* Annals of the American Academy of Political Science, No. 65, 1916, pp. 195-204.



Graph I. Ratio of Median M.A. to Median Ch. A.

chronological age. If the median mental age with advance in grade increases faster than the median chronological age, then the difficulty of passing each successive grade is probably on the increase. Inasmuch as an increase in the ratio used as a measure of this may be due to the nature of the units used, we should compare the *rates* of increase of the curves rather than the absolute amounts of increase. We see that School A, of highest status, has a consistently greater rate of increase than School C, the lowest in social status, and which is taken by way of affording a good contrast. Accordingly, since it is harder in School A for a child of normal intelligence to pass each successive grade than for a similar child in School C, we should expect School A to have much the larger percentage of retardation. Actually, however, School A has a smaller percentage, 33.5 for the entire school as compared with 38.4 similarly for School C. A possible explanation, although one on which we have no data, is that perhaps School A eliminates a larger percentage of those who lag behind the normal rate of progress. In School A we find much smaller percentages of very badly retarded pupils than in School C. School A consistently has the lowest Q-values of the three schools, such as we might be led to expect where the standards are changing the most rapidly from grade to grade. Whether increasing standards and a lower Q-value are desirable, hinges upon a question of school policy, "Shall we have high standards and fewer 'educated' pupils, or shall we have lower standards more nearly perfectly suited to those at the lower end of the mentality scale and a larger number of 'educated' pupils!"

Summary

1. The association coefficient, Q, for the relation of the four classes, School Retarded and Below Median in Mentality, School Retarded and Above Median in Mentality, School Advanced and Below Median in Mentality, School Advanced and Above Median in Mentality, was found to be +.85.

2. Q, for the classes, School Retarded and Mentally Backward or Below, School Retarded and Mentally Bright or Above, School Advanced and Mentally Backward or Below, School Advanced and Mentally Bright or Above, was found to be +.97.

3. Four grades or more retarded in school is a perfect indication of a mentality not higher than normal, with the chances greatly in favor of its indicating at least a backward mentality. Two grades or more advanced in school are likewise a perfect indication of at least a normal mentality, with the chances greatly in favor of its indicating at least a bright diagnosis.

4. Wherever low indices are found, the most important reason for this is the tendency of the schools to retard children who are perfectly normal or slightly advanced mentally.

5. If this tendency is explained by lack of industry on the part of pupils who would presumably be promoted regularly with industry present, then a possible cause for the lack of zeal of these normal pupils is the increasing difficulty of passing the more advanced grades. This is a school tendency which is most prominent in the school of highest social status.

6. Whether increasing standards are desirable is a matter of school policy, to be determined by the educational aims recognized by the different schools. If the aim is the selection and fullest education of the better mentally endowed children rather than graduation of the largest possible number of pupils, then a high value for Q is undesirable.

THE EFFECTS OF SPECIAL DRILL IN ARITHMETIC AS MEASURED BY THE WOODY AND THE COURTIS ARITHMETIC TESTS

J. E. EVANS AND FLORENCE E. KNOCHE

Ohio State University

Introduction.—Numerous and important factors are entering into the public school problem of today, which emphasize the need of efficient, time-saving, and systematic methods of drill work, so necessary in the formative period of our school youth. The question arises: Is a systematic form of drill work a saving over the regular school work, and is it just as much or more efficient?

Plan of the Test.—The following data, charts, and results in arithmetic are based upon the work of a 6A and a 6B grade in a Columbus City School. The work extended over a period of 43 school days from October 4, 1917 to January 29, 1918. The 6B group was used as the practice group and the 6A as the control group. The Courtis Test and the Woody Arithmetic Test A were used for control tests. The Studebaker Economy Practice Exercises* were used as the daily drill work in addition to the regular school work in arithmetic, except the time of the drill was deducted from the regular arithmetic work of the 6B class. The 6A class received no training in arithmetic except the regular school arithmetic. Both grades were tested with the Courtis, and Woody Arithmetic tests in October and again in January after the 43 periods of practice.

Personnel of the Class.—All the children from two rooms were selected for the experiment. Each room was in charge of the regular teacher and no attempt was made to make the slight change in the program due to the introduction of the drill as something special, in fact, no child knew that the work was not a part of the regular prescribed work for the semester. The 6B class was in charge of one of the writers. The children in the two rooms were quite similar in ability. The 6A class was one semester in advance of the 6B group. The median age of the children in the 6A group was 11.7 years, and 11.5 years for the 6B group. During the period of drill the main work for 6A grade was percentage with a general review of the fundamental processes. The work of the 6B grade was decimals. As one of the writers has had experience in teaching both grades she believes that both grades were average ability in their respective classes.

---

*STUDEBAKER, J. W. *Teachers' Manuals for the Studebaker Economy Practice Exercises in Arithmetic.* p. 2. (Sott, Foreman & Co., N. Y.)

Description of the Control Tests.—(A) The Courtis Standard Research Tests, Series B, Form 1, are a series of problems in addition, subtraction, multiplication, and division. The problems represent regular formal problems in the various processes and one problem in a list is about as difficult as any other problem in the same list. The emphasis is on speed and accuracy. Only a knowledge of the four fundamental processes is necessary in order to solve the problems. A time limit is set for each process. The score records the number of problems attempted and the number solved correctly. Thus the score made by the class can be compared with a previous score, or the form set up by Mr. Courtis.

(B) The Woody Arithmetic Scale was also used in the control series. This series consists of four scales, one for each of the four fundamental processes. Woody states that his "fundamental idea was to derive a series of scales which would indicate the difficulty of the problems that a class can solve correctly. Accordingly, each of the scales is composed of as great a variety of problems as the fundamental operations can well permit.[1] These problems, beginning with the easiest that can be found, gradually increase in difficulty until the last ones in each series are so difficult that only a relatively small percentage of the pupils of the eighth grade are able to solve them correctly.[2]

A time limit of 20 minutes was allowed for each process. This allows ample time for even the slowest pupil to solve all the problems which he is capable of solving. This method, while it places the emphasis on achievement and does not consider speed as an important factor in ability, is a good test of the regular class work, yet it must be remembered that the courses of study throughout the country are not uniform. In the Columbus Course of Study little emphasis is placed upon denominate numbers in the sixth grade. As a result these pupils were unable to solve problems of this type. The four tests contain 10 such problems.

Description of the Drill Work.—The Studebaker Economy Practice Exercises (SET B) include 50 different exercises and are to be used in grades four to eight inclusive. There are 16 exercises in addition, 7 in subtraction, 13 in multiplication, and 14 in division. The amount of material in the set is sufficient for a year's work, although some very bright children will complete all the exercises in a much

[1]Written problems are not included by Mr. Woody.

[2]Woody, Clifford. *Measurements of some Achievements in Arithmetic.* Columbia Contributions to Education, No. 80, 1916, p. 1.

shorter time. The examples have been carefully formulated according to the author, and the graded standards are based on the tests given to several thousand pupils.

As the child works through the exercises he will review in an orderly way all the possible combinations in the four fundamental processes. Increased skill is secured not by making the examples more difficult, but by requiring pupils in the higher grades to do the work in less time. For example, the fourth grade children of ordinary ability are expected to complete any one of the exercises successfully in 7 minutes, whereas the eighth grade pupils should do the same work in four minutes.

The following is Studebaker's Table of Standards:

Number of minutes for completion with 100% accuracy	Grade				
Classes	4	5	6	7	8
Classes of ordinary ability	7	6	5	4½	4
Classes of exceptionally capable children	6½	5½	4½	4	3½
Classes of exceptionally slow children	7½	6½	5½	5	4½ *

The graphic representation of the difference in achievement in the two groups is shown in Plates II and III. This graph also gives a comparison of the achievement of the class in October with the achievement of January. The graphic representation of the norm of Courtis is also included in the graph for comparison purposes.

The Efficiency of the two Groups at the End of the Practice Period.—The degree to which teaching increases the efficiency of the class is very important. Mr. Courtis believes that the teaching efficiency of arithmetic as taught at present is about 5% in the United States. A not impossible ideal in such a subject as arithmetic would be 100%, but 80% would be a fair achievement if absolute accuracy is not to be the goal. For want of a better statement of representation of efficiency it will mean here the percentage of the class which reaches or exceeds the norm (median) for the group. For the sixth grade in addition it will mean the percentage of the class which solves 9 or more problems correctly in 8 minutes in the Courtis Arithmetic list.

*Studebaker, J. W. *Teachers' Manuals for the Studebaker Economy Practice Exercises in Arithmetic.* p. 2. (Sott, Foreman & Co., N. Y.)

PLATE I.
THE STUDEBAKER ECONOMY EXERCISE

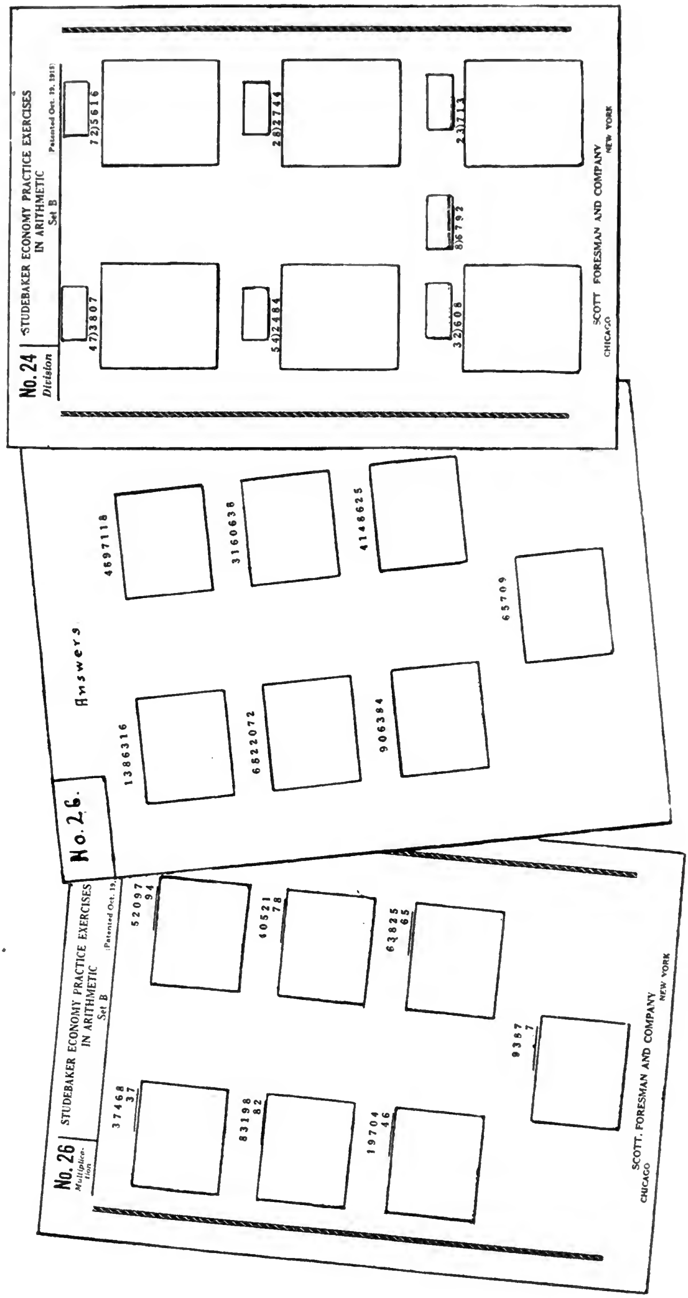


TABLE I.

Results of the Drill.—The percentage of gain from October to January by the Woody Arithmetic Test is shown in Table I.*

Process	Grade	Date			Boys Aver. Accu.	Girls Aver. Accu.	Total Aver. Accu.	Total Aver. Gain	Gain % 6B over 6A
Addition	6A	Oct. 8	1917		69.9%	67.7%	68.8%		
		Jan.28	1918		74.4	82.8	78.6		
				Gain	4.5	15.1		9.8%	
Addition	6B	Oct 8	1917		71.2	68.9	70.1		
		Jan. 28	1918		85.8	86.7	86.3		
				Gain	14.6	17.8		16.2	+6.4
Subtraction	6A	Oct. 8	1917		74.1	68.3	71.2		
		Jan.28	1918		72.5	72.7	72.6		
				Gain	-1.6	+4.4		1.4	
Subtraction	6B	Oct. 8	1917		72.2	63.9	68.1		
		Jan.28	1918		80.0	78.5	79.3		
				Gain	7.8	14.6		11.2	+9.8
Multiplication	6A	Oct. 8	1917		56.6	55.7	56.2		
		Jan.28	1918		72.2	71.0	71.6		
				Gain	15.6	15.3		15.5	
Multiplication	6B	Oct. 8	1917		58.3	58.8	58.6		
		Jan.28	1918		65.7	69.0	67.4		
				Gain	8.4	10.2		9.3	—6.2
Division	6A	Oct. 8	1917		60.5	61.1	60.8		
		Jan. 28	1918		70.2	72.7	71.5		
				Gain	9.7	11.6		10.07	
Division	6B	Oct. 8	1917		59.2	58.7	69.1		
		Jan.28	1918		79.0	72.0	75.5		
				Gain	19.8	13.3		16.5	+5.8

TABLE II.

A comparison of the achievement of the two groups with the Woody norm.

Process	Grade	Date		Class norm	Woody norm	Gain
Addition	6B	Oct.	1917	7.72	7.95	
		Jan.	1918	8.59		.87
	6A	Oct.		7.62	7.95	
		Jan.		7.91		.29
Subtraction	6B	Oct.		6.27	6.46	
		Jan.		7.51		1.24
	6A	Oct.		6.17	6.46	
		Jan.		7.31		1.14
Multiplication	6B	Oct.		6.21	6.72	
		Jan.		6.82		.61
	6A	Oct.		6.19	6.72	
		Jan.		6.98		.79
Division	6B	Oct.		5.70	5.87	
		Jan.		6.56		.86
	6A	Oct.		5.89	5.87	
		Jan.		6.55		.66

*In Table I the percentage of accuracy is the percentage of problems solved correctly of all the problems attempted and not of the entire number in the list.

SUMMARY OF TABLE I. *

Process	Gain of 6B over 6A
Addition	6.4%
Subtraction	9.8
Multiplication	-6.2
Division	5.8
Average Gain	5.8

* In Table I the percentage of accuracy is the percentage of problems solved correctly of all the problems attempted and not of the entire number in the list.

TABLE III.

The percentage of gain from October to January as shown by the Courtis Arithmetic Test.

Process	Grade	Date		Boys	Girls	Aver.	Gain	Gain of 6B over 6A
Addition	6A	Jan.	1918	54.9+	60.8	57.9		
		Oct.	1917	51.8	52.5	52.2	5.7%	
	6B	Jan.	1918	74.5	81.6	78.1		
		Oct.	1917	69.1	52.1	60.6	17.5	11.8%
Subtraction	6A	Jan.	1918	73.8	64.4	69.1		
		Oct.	1917	80.8	69.5	75.2	-6.1	
	6B	Jan.	1918	98.1	86.7	92.4		
		Oct.	1917	76.7	71.0	74.4	18.1	24.1%
Multiplication	6A	Jan.	1918	55.2	68.2	61.7		
		Oct.	1917	57.4	56.0	56.7	5.0	
	6B	Jan.	1918	74.0	78.0	76.0		
		Oct.	1917	68.7	59.0	63.9	12.2	7.2%
Division	6A	Jan.	1918	65.2	71.2	68.2		
		Oct.	1917	64.2	63.7	64.0	4.3	
	6B	Jan.	1918	78.8	71.6	75.2		
		Oct.	1917	56.2	61.5	58.9	16.4	12.1%

SUMMARY OF TABLE III.

	6A		6B	
Addition	5.7%	gain	17.5%	gain
Subtraction	-6.1		18.1	
Multiplication	5.0		12.2	
Division	4.3		16.4	
Average Gain	2.2		16.1	

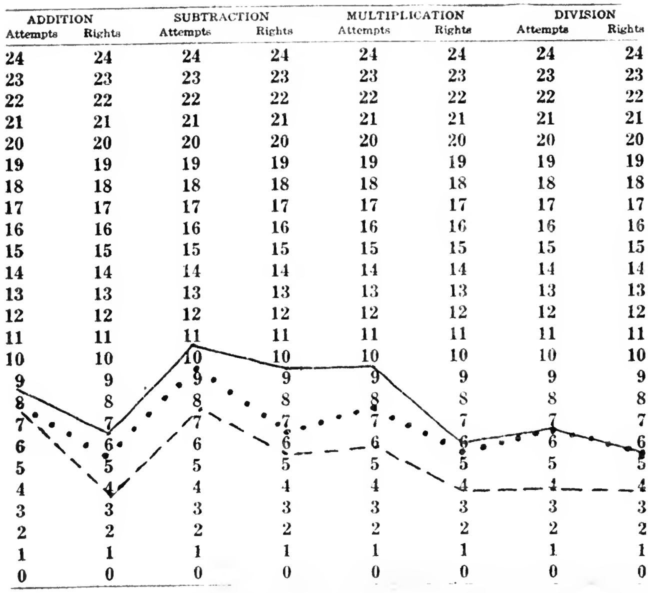
PLATE II.

Class Score

6B

Test	Subject	Attempts		Rights	
No. 1.	Addition	7.1	8.3	3.8	6.6
No. 2.	Subtraction	7.1	10.3	5.7	9 5
No. 3.	Multiplication	5.8	9.3	3.6	6.0
No. 4.	Division	3.8	6.5	3.5	5.3
		Oct.	Jan.	Oct.	Jan.

GRAPH.



The— — — —— line indicates the graph of the , school for October
The ———————— line indicates the graph of the school. for January
The . . . . line indicates the standard graph by Courtis.

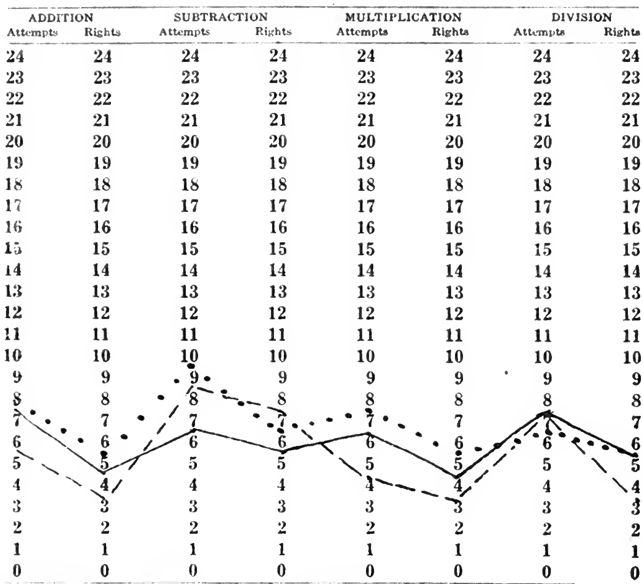
PLATE III.

Class Score

6A

Test.		Subject	Attempts		Rights	
No.	1.	Addition	5.6	7.1	3.2	4.6
No.	2.	Subtraction	8.9	6.9	7.6	5.8
No.	3.	Multiplicat'n	4.1	6.7	3.7	4.6
No.	4	Division	7.5	7.3	3.3	5.3
			Oct.	Jan	Oct	Jan

GRAPH.



The — — — — line indicates the graph of the school for October.
The ———————— line indicates the graph of the school. for January
The . . . . . . line indicates the standard graph by Courtis.

PLATE IV A.

6 A.— October

Test No. 1. Addition. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100				1		1																				2	6%
90																										0	0%
80							2			1																3	9%
70					2						1															3	9%
60				1		2			1	2																6	18%
50			1		1		1																			3	9%
0 to 49			1		4	5	2	3	1																	16	48%
Total			2	2	7	8	5	3	2	3	1															33	

Median Scores: Speed 5.6 Accuracy (23) 57.1% Efficiency 0.0% Standard Scores 9-9

Test No. 2. Subtraction. Standard Scores: Grade 3, 4; Gd 4, 6; Gd 5, 8; Gd 6, 10; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100				1	3	2	3			1																10	30%
90												1														1	3%
80						2	2				2															6	18%
70					4				1	1																6	18%
60						2																				2	6%
50												1														1	3%
0 to 49					1	2	3						1													7	21%
Total				1	8	8	8	0	1	2	2	2	1													33	

Median Scores: Speed 7.9 Accuracy (7.6) 83.4% Efficiency 0.0% Standard Scores 10-10

For instructions, see other side of this page.

6. A. — January

Test No. 1. Addition. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100																										0	0%
90											1															1	3%
80						3	2		1																	6	18%
70								3	1	1	1	1														7	20%
60						1	2		1																	4	11%
50					3			2	1		1															7	20%
0 to 49					3	3	2	1																		9	23%
Total					6	7	6	6	4	1	3	1														34	

Median Scores: Speed 7.2 Accuracy (44) 64.8% Efficiency 0.0% Standard Scores 9-9

Test No. 2. Subtraction. Standard Scores: Grade 3, 4; Gd 4, 6; Gd 5, 8; Gd 6, 10; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100								1		2																3	9%
90																										0	0%
80						1	2		4	2																9	23%
70					2		2	1		2																7	20%
60						5	2					2														9	23%
50						1	1																			2	6%
0 to 49						1	2	1																		4	11%
Total					2	8	9	3	4	6	0	2														34	

Median Scores: Speed 6.9 Accuracy (5.8) 84% Efficiency 0.0% Standard Scores 10-10

PLATE IV B.

Test No. 3. Multiplication. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100				1			2																			3	9%
90																										0	0%
80						1	2			1	1	1														6	18%
70					2			2	1	1																6	18%
60						1	1																			2	6%
50			1		2		1	1																		5	15%
0 to 49				4		4	1	2																		11	33%
Total			1	5	4	6	7	5	1	2	1	1														33	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 4.1 Accuracy (3.7) 90.2% Efficiency 0.0 % Standard Scores 9—9

Test No. 3. Multiplication. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100					1		2	1	1	1																6	18%
90													1													1	3%
80						1	2																			3	9%
70										1		1														2	6%
60							3		2	1																6	18%
50					1		3	2																		6	18%
0 to 49				1	2	5	1	1																		10	29%
Total				1	4	6	11	4	3	3	0	1	1													34	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 6.7 Accuracy (4.6) 68.6% Efficiency 2.9% Standard Scores 9—9

Test No. 4. Division. Standard Scores: Grade 3, 2; Gd 4, 4; Gd 5, 6; Gd 6, 8; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100			1	3	4			1																		9	27%
90											1															1	3%
80							1	1	1																	3	9%
70					4			2																		6	18%
60				2			1																			3	9%
50			1		3																					4	12%
0 to 49			1	1	4	1																				7	21%
Total																										33	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 7.5 Accuracy (3.3) 44% Efficiency 0.0 % Standard Scores 8—8

6A.—October

Test No. 4. Division. Standard Scores: Grade 3, 2; Gd 4, 4; Gd 5, 6; Gd 6, 8; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100								1	1	1		1		1												5	14%
90												1														1	3%
80							2	3		1																6	18%
70					1			2				1														4	11%
60						2	3		1		1		1													8	21%
50							2	1																		3	9%
0 to 49				1	1	2	1	2																		7	20%
Total				1	2	4	8	9	2	2	1	3	1	1												34	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 7.3 Accuracy (5.3) 72.6% Efficiency 11.7% Standard Scores 8—8

6·A—January

PLATE V A.

6 B.— October.

Test No. 1. Addition. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6. 9; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100																										0	0%
90									1			1														2	5%
80								1	1	2	1															5	13%
70					1		1			1	1															4	11%
60				1		1	3	2		1																8	21%
50					2	1	2		1																	6	16%
0 to 49			1		3	1	2	2	2	1		1														13	34%
Total			1	1	6	3	8	5	5	5	2	2														38	

Accuracy. Figures in small type indicate scores in number of examples right.

Median Scores: Speed 7.1 Accuracy (3.8) 53.5 Efficiency 0.0% Standard Scores 9—9

For instructions, see other side of this page.

Test No. 2. Subtraction. Standard Scores: Grade 3, 4; Gd 4, 6; Gd 5, 8; Gd 6, 10; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100				1	1	1	1			3																7	18%
90																										0	0%
80						3	2		4	1	1															11	29%
70					3			1	2		2															8	21%
60				1		1	1				1															4	11%
50							2	2																		4	11%
0 to 49				1			1	2																		4	11%
Total				3	4	4	8	5	6	4	4															38	

Accuracy. Figures in small type indicate scores in number of examples right.

Median Scores: Speed 7.1 Accuracy (5.9) 71.8% Efficiency 0.0% Standard Scores 10-10

For instructions, see other side of this page.

6 B.— January.

Test No. 1. [redacted] Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100					1	1				2	1	1			1											7	21%
90													1													1	3%
80						1	2	3	1	2	1			1												11	33%
70					1				1	1			1													4	12%
60						1	2			1	1															5	15%
50					2		1	1			1															5	15%
0 to 49																											
Total					4	3	5	4	2	6	4	1	2	1	1											33	

Accuracy. Figures in small type indicate scores in number of examples right.

Median Scores: Speed 8.3 Accuracy (6.6) 79.5% Efficiency 15.1% Standard Scores 9-9

Test No. 2. Subtraction. Standard Scores: Grade 3, 4; Gd 4, 6; Gd 5, 8; Gd 6, 10; Gd 7, 11; Gd 8, 12. Accuracy 100%

Score in number of Examples Attempted

Score	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
% 100						2		1	1	1	5				1											11	33%
90											2	1	4	1			1									9	27%
80						1		2	2	2	1		1	1												10	30%
70									2			1														3	9%
60																											
50																											
0 to 49																											
Total						3	0	3	5	3	8	2	5	2	1	0	1									33	

Accuracy. Figures in small type indicate scores in number of examples right.

Median Scores: Speed 10.3 Accuracy (9.5) 93.2% Efficiency 18.1% Standard Scores 10—10

PLATE V B.

Test No. 3. Multiplication. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy %)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
100				2	1	1		2																		6	10%
90																										0	0%
80						1	2	3	1																	7	18%
70					2			3	1																	6	10%
60						1	2		1																	4	11%
50			1		4																					5	13%
0 to 49		2			2	4	1		1																	10	26%
Total		2	1	2	9	7	5	8	4																	38	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 5.8 Accuracy (3.6) 12.0% Efficiency 0.0% Standard Scores 9-9

Test No. 3. Multiplication. Standard Scores: Grade 3, 3; Gd 4, 5; Gd 5, 7; Gd 6, 9; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy %)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
100					1	1	1	1		·		1			1		1									7	21%
90												1														1	3%
80						1		1	2	1			2	1		2										10	30%
70								2	1	1			1		1											6	18%
60						1	1			1						1										4	12%
50					1			1	1	1		1														5	15%
0 to 49																											
Total					2	3	2	5	4	4	0	3	3	1	2	3	1									33	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 9.3 Accuracy (60) 44.5% Efficiency 9.1% Standard Scores 9-9

Test No. 4. Division. Standard Scores: Grade 3, 2; Gd 4, 4; Gd 5, 6; Gd 6, 8; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy %)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
100			2	7	2				1	1																13	34%
90																										0	0%
80						2	1	1																		4	10%
70					2																					2	5%
60				3		1	1																			5	13%
50			2		1																					3	7%
0 to 49		2	2	3	2	1				1																11	28%
Total		2	6	13	7	4	2	1	1	2																38	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 3.9 Accuracy (3.5) 89.7% Efficiency 5.2% Standard Scores 8-8

6 B.—October

Test No. 4. Division. Standard Scores: Grade 3, 2; Gd 4, 4; Gd 5, 6; Gd 6, 8; Gd 7, 10; Gd 8, 11. Accuracy 100%

Score in number of Examples Attempted

Score (Accuracy %)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	Total	%
100		1	1	1	1	2	1	1	1		1		2													12	36%
90											1															1	3%
80						3	2	1	2	3																11	33%
70					1			1																		2	6%
60				1						1																2	6%
50			1		1		1																			3	9%
0 to 49				2																						2	6%
Total		1	2	4	3	5	4	3	3	4	2	0	2													33	

Figures in small type indicate scores in number of examples right.

Median Scores: Speed 6.5 Accuracy (53) 81.5% Efficiency 12.1% Standard Scores 8-8

6 B.—January

Tables III and IV show the efficiency and also the distribution of the two groups in October and in January. The percentage of the class which falls above and to the right of the heavy black line will represent the efficiency of the group. It is easy to note also that the distribution of the 6A group is about the same in January as in October, but the position of the 6B grade has changed completely.

It would be quite interesting to make such a comparison with the Woody scale but the type of norm used prevents individual comparisons with the norm. The achievement of the two groups may be compared, however, with the Woody norms. This comparison is shown in Table V. In every case except one (Multiplication) the 6B group made a greater gain than the 6A group. The teacher felt that even this one exception was due to the special emphasis which she had placed upon this process in the review work during the period of practice.

The Effect of Practice upon the Bright and the Dull Members of the Class.—A very valid question to ask concerning every mechanical device or drill of this nature is whether the poorest or slowest pupils of the class who really need the assistance most are really benefited as much as the faster and brighter members of the class. To answer this question the 6B group was divided into two parts according to the number of exercises which the members of the class solved or worked with 100% accuracy in the drill. The set contained 50 exercises but the greatest number solved in the 43 days was twenty-five. Those pupils who solved twenty-five exercises were termed the fast or bright group. The slow or dull pupils include those who solved eight exercises or less in the same time. The first group gained 17% in accuracy in addition in the Woody exercises, and 24% in the Courtis exercises. The slow group made a gain of 15% and 31% respectively in the same time. What is true of addition was also true of the other processes. It is interesting to note that in the two groups there was not a single case in which the individuals were not benefitted by the exercise or drill.

That the drill was as effective for the older members of the class as the younger members is shown in Plate V in the graphic representation for each individual for the four processes in the Courtis test both in the October and the January examination. The unshaded oblongs represent the achievement in October, and the

shaded oblongs the achievement in January. The ordinates show the number of problems solved correctly; the abscissa show the number of cases arranged in an increasing series according to the age of the individual as shown by the figures along the base. In some few cases the child was not present for both tests. In the sixth grade children of 13 or 14 years of age are apt to be retarded from some cause. A weak mentality is not the least among these causes. In this class some few cases may have been retarded because of low mentality, but in all these cases the drill was effective, nevertheless.

Conclusions.—Such drill exercises as the one here used have a great value in increasing the ability of a class in terms of accuracy far beyond the class exercises and drills now given in the public schools for the same purpose. This special exercise benefitted all members of the class regardless of whether they were slow or fast, dull or bright, very young or old, for the sixth grade. The average increase in efficiency due to 43 periods of practice of 5 minutes each (the maximum time used in preparation for the drill, and the drill, was 10 minutes) was 17.3% for the four processes. The increase in accuracy for the class with only the regular school work was 3.7%.

Both classes were below the Courtis and Woody norms in October. In January the class with the special drill had reached the norm for division in the Courtis and had surpassed the norms in both scales in the other cases. The class without the drill was still below the Courtis norm in all except division, and below in addition in the Woody scale.

As a convenient, effective and time-saving method of gaining speed and accuracy in the fundamental processes in contrast to the regular work of the school the drill is quite effective. The drill seems to be the most effective for the type of mathematical ability measured by the Courtis test rather than the ability which the Woody scales measure, although all the pupils made gains in this direction as well.

WORK CURVES

THOMAS RUSSELL GARTH

West Texas State Normal College, Canyon, Texas

Meumann, Kraepelin and others claim that there are different types of work-curves in computation experiments. Meumann claims that there are three types of workers as indicated by their curves. One of these obtains maximal efficiency at the start and decreases with many fluctuations; the second obtains maximal efficiency only after an interval of great length. The first is characterized by quick adaptation and early fatigue; the second by slower adaptation and slower fatigue; and, the third, by very slow adaptation and high resistance to fatigue.*

Offner speaks of a fourth type, the "concave type of curves." He says then: "It is possible to regard the peculiarities in the general shape of the work-curve as due simply to individual differences in type of fatigue; in that case, the curves would be really curves of fatigue." †

The writer in his discussion of individual curves ‡ calls attention to the fact that in the data examined, there are various kinds of curves. Such differences may be due to slight intercurrent causes, and are not, in the lack of further evidence, to be regarded as indicative of different types of workers. And since the data offered peculiar opportunity to show whether or not there are such things as types in computation or addition work-curves, the material of the above mentioned experiment has been made use of for that purpose and the results are given in the following pages.

Description of Material to be Handled

The material to be handled is data obtained in a continuous work experiment in which three hundred sixty-eight third and fourth grade, and three hundred forty-three seventh and eighth grade children of the Virginia public schools, — Richmond, Petersburg, and Farmville used the Thorndike addition sheets. The purpose of the original experiment was to find out how much school children

*G. M. Whipple. *Manual of Mental and Physical Tests.* p. 335.

† M. Offner. *Mental Fatigue.* pp. 75-76.

‡ T. R. Garth. *Mental Fatigue During Continuous Exercise of a Single Function.* p. 26. Archives of Psychology, Vol. XXVI, No. 2 Columbia University Contributions to Philosophy and Psychology.

fatigue while engaged in doing addition during a rather extended length of time. The addition sheets were arranged in pads securely fastened together. Before the experiment was begun these pads were passed out to the pupils face downward and each child was instructed to write his name, school, grade, the date and hour, and his age and sex on the blank side. The explanation was made to the pupils that the experimenter wished to find out who was the best "adder" in the class and it was insisted that each should do his best. In addition they were told that they would spend two minutes on a single sheet as timed by the experimenter's stop watch. They were to have the signal "Go!" for starting together on a sheet, and "Stop!" for a signal that the two minutes were up. No time was lost in passing to the next sheet for the experimenter said immediately, "Take the next sheet. Go!" This procedure continued until all the sheets in the pad were used up.

When the sheets were taken up and gone over the experimenter was able to know just how much was done by each child every two minutes.

Many of them did better at the start than they did toward the finish, as one would naturally suppose, but some did better at the close of the working time than they did in the beginning. However there was more of a tendency toward falling away in accurate work at the end than there was in the merely attempted work. That is to say that a pupil undertook more problems than he did correctly.

The group performances were plotted into curves whose points were determined by the average performance of the group at each two minutes and the study of these curves show a falling away at the end when it is taken in comparison with the beginning of the curve. The maximum of accurate performance as indicated by the curves is reached earlier in the course of the work than in the attempted performance.

Space is lacking here to describe the experiment more fully, but the outcome of the experiment, speaking briefly, was that the younger group experienced a greater falling off in both work attempted and in accurate work that the older group. Other conclusions besides were drawn, such as that the slow workers tended to fatigue sooner than the better workers, but these observations are not necessary to the discussion here in hand and they are only given to afford the reader some idea of the nature of the original experiment.

The younger group added for twenty-eight minutes and the older group for forty-two minutes. The records show what was attempted and what was accurately done by each subject for each two minutes of the time. We call these two-minute divisions of the total time, *periods*, making fourteen periods for the younger group and twenty-one periods for the older group.

In this way, we get three hundred sixty-eight curves of work attempted, and three hundred sixty-eight curves of accurate performance in the lower grades; and in the higher grades we get three hundred forty-three curves of attempted and three hundred forty-three curves of accurate performance. The total amount attempted and accurately performed by each child we now ascertained for every single curve. Because it was felt in making a group curve of the absolute data, as above described, that the poorer workers were placed at a disadvantage and that their individual curves of work did not obtain a fair representation; in addition to the absolute curves, another kind of curve for each rubric (attempts and accurates), for each group, was made. This curve is called the average of the individual curves. To obtain the curve, the curve of each child was expressed in per cents of the total performance and, consequently, each step in the curve shows what part of his total work was done by him at that period of the total time allotted by the experiment. This seemed to give a more adequate expression of the work of the group than a curve expressed in absolute measures, or columns of addition. This average of the individual curves is somewhat different from the absolute curves.*

The curves just mentioned, composed of the per cent. of total performance done in each period, are the material upon which we wish to base these observations and conclusions as to types of work-curves.

The question as to whether there are types of work-curves in the same sort of work, as in addition of one place figures, may be answered by the generalization that there are no such things as many types in measurements of any one kind of human process or trait, and that there is only one type. This is a brief statement of the "Single-Type Theory." The alternative to this would be the "Multiple-Type Theory." If the Single-Type Theory is true, then there are no such things as types of work-curves, and Meumann and Kraepelin are wrong.

*A discussion of this may be found on pages 9-14 of the writer's dissertation, *Mental Fatigue During the Continuous Exercise of a Single Function*. Archives of Psychology, Columbia University.

In order to let these data contribute their experimental light on the subject, we have arrayed them in a way which should make of them a more or less positive argument for the support of the Single-Type Theory.

Work- Curves

Let us consider that we have all the individual curves arrayed before us. It is true, that some of the workers do rather well at the start and rather poorly at the end, some somewhat poorly at the start and somewhat better at the end, and others do their best work in the middle of the total time allotment. These curves may be found scattered all along, in going from the poorest performances on up to the best ones. But it would be another thing to call these types. We can prove this one way or the other, only by placing the data of the materials for the different periods or groups of periods in distribution curves to see if they are unimodal or not. If there are three types of work-curves, we should have three modes. If there is only one mode or central tendency of the facts, then there is only one type of work-curve.

Methods Used

As suggested above, we may examine the distribution belonging to the averages, which go to make up the curve representing the so-called average of the individual curves at any point along the self-same curve. Or better, to avoid chance errors, we may take into consideration how much was done during the first three periods by each individual and arrange these facts in a distribution curve. If there were three types, two would be evident at the start, and we should have then a bimodal distribution there. In order to further test the data for types, we could take the sum of what was done in the first seven periods, or half the total time of working in the younger group, and in the first eleven periods by the older children, and making distributions of these. Multiple modes ought to be evident here, two at least, if there are types of work-curves. Again, we could distribute the sums of what was done in the last three periods. If there are multiple types of work-curves, it should be evident here by bi-modality at least.

We give herewith the curves representing the distribution of these facts, those of the third and fourth grades first, followed by those of the seventh and eighth grades.

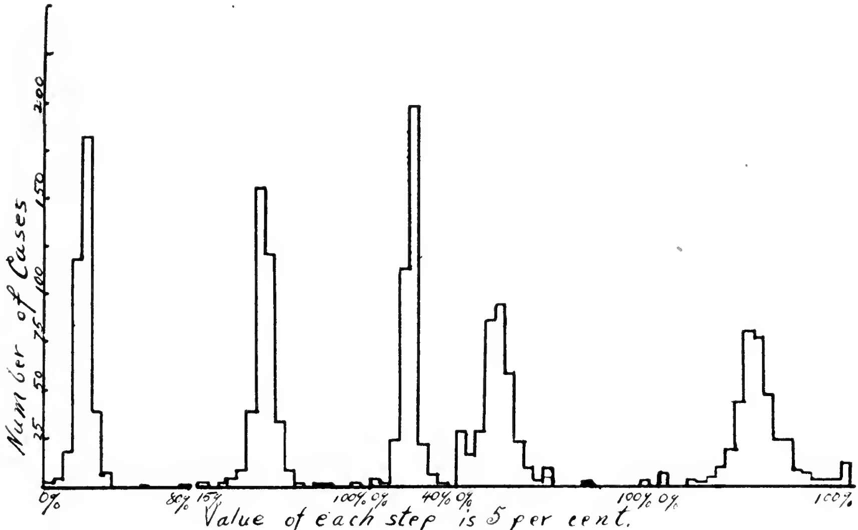


Fig. 1 Fig. 2 Fig. 3 Fig. 4 Fig. 5

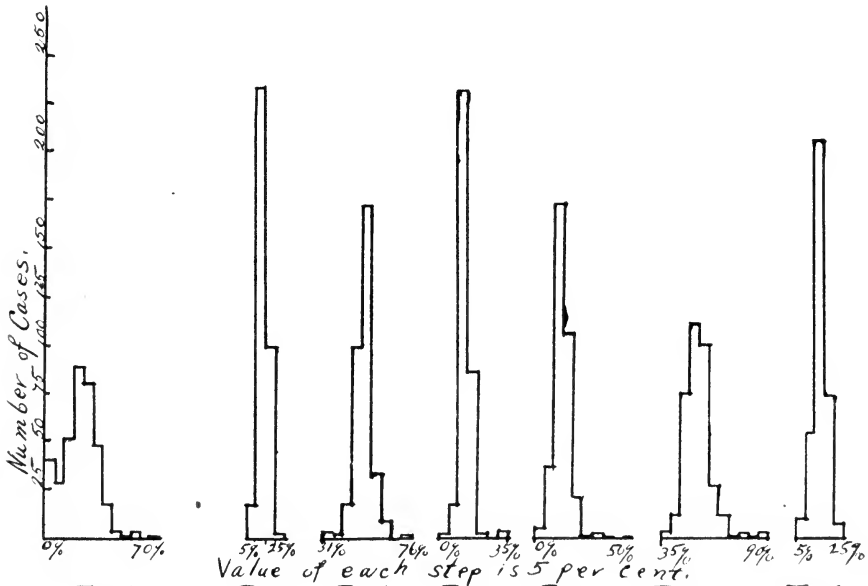


Fig. 6 Fig. 7 Fig. 8 Fig. 9 Fig. 10 Fig. 11 Fig. 12

The Distribution Surfaces

First, tables of the distributions were made. These tables were then condensed into a coarse grouping. If there is any real tendency to multimodality in the distributions, it should make itself known by this severe test of the coarse grouping, and even if it appears in this case it may not be of such a nature as to argue for many types of work-curves in the data. Figures 1 to 12 present the facts of the coarse groupings in distribution curves.

These curves show in more or less strong degree, a tendency to take the form of the normal distribution curve, with seeming departures therefrom in Figures 4 to 6. The tendency to multimodality appears at the beginning of all these last mentioned curves and at the end of one of the three — Figure 5.

The detailed analysis of the facts represented by these points of the curves must be made so as to determine whether or not types are here indicated by the distribution surfaces. It will be well to note the curves carefully, so as to follow the discussion. In all these curves, the frequency for the 0 to 5 step stands somewhat above that of the 5 to 10 step. In Figure 5, the frequency of the 95 to 100 step is some greater than that of the 90 to 95 step.

These curves represent the accurate performance of the younger group of children, the third and fourth grade groups, and the secondary modes represent the work of the poorer performers. If it had been possible to employ a finer measure of the work accurately done, these departures from the normal surface would not be here.

In Figure 4 the individuals represented by a frequency of 27 at the 0 to 5 step, had a total correctly added ranging from 0 to 12 columns, viz: 19 pupils had a correct score of 0 to 6; 8 pupils had 6 to 10 correct; 3 pupils had 11 to 12; and, 1 had 19 correctly added columns. When a pupil added only one column correctly duiing the whole fourteen periods during which he worked, his per cent. would be 100 for that period. No matter how nearly accurate he may have been at any other time, just this much of accurate performance had to be ignored because we had no way of saving it or even knowing its amount, for a subject was credited with nothing less than one correctly added column. Consequently, zero does not represent all of one's work; as Thorndike says, the zero point does not always represent absolute zero but is often an indefinite point on the scale. In a situation such as the one here in hand, we have what he calls an "undistributed or indefinite" measure. He

furthermore says, "One should always guard against undistributed measures at either extreme of the scale."* The situation then is due to inadequate measurement, here and in the other curves, at the zero to five step. In the distribution represented by Figure 5, the same source of error plays at the 95 to 100 step when a frequency of 12 shows that a like number (12) of pupils did from 95 to 100 per cent. of their total work during the first-half of the total time of the experiment.

Another source of error of which we must take account† in such measurements as these data represent is that error which is due to a small number of cases which gives rise not only to apparent mutimodality, but to gaps in the distribution surface. In many of the twelve curves given herewith, we find illustration of this fact which a larger number of subjects would no doubt have altered.

CONCLUSION

We must conclude, therefore, that if our data could be corrected for these errors, that the curves would show without exception, conformity to the normal distribution curve. If we accept this inference, that is, if there are no types to be found in the distributions whose several averages have been taken consecutively, as composing the average of the individual curves, then there is found in this data no evidence whatever of types of work-curves, but a tendency on the part of the individual workers to follow the average of the individual curves. And this conclusion, if it may be taken as an inference satisfactorily drawn from these premises, is, within the limits of the meanings of these data, but another argument to pursuade us to believe with Thorndike, that "A multitude of nearly equal factors from which each man's nature and training is approximately a random selection, will always act towards the production of unimodality" and "that in proportion as exact measurements have been applied, evidence expected to favor the Multiple-Type Theory has turned out in favor of the Single-Type Theory."‡

‡E. L. Thorndike. *Mental and Social Measurements.* p. 22.

†E. L. Thorndike. *Educational Psychology, Vol. III.* p. 34.

‡E. L. Thorndike. *Educational Psychology, Vol. III,* p. 345 and p. 379.

STANDARDIZED ILLUSTRATIVE SENTENCES FOR THE SPRINGFIELD SPELLING LIST

DAVID SPENCE HILL

The University of New Mexico

In the administration of standard tests in spelling where words are dictated, it is customary for the examiner or teacher to illustrate the meaning of a given word immediately after uttering it. It seems obvious that the response of the child will be affected in minor or major degree by the character of illustrative sentence used by the teacher. Ordinarily, a simple and felicitous sentence will be used, but on the contrary, there are occasions where the extemporized sentence obscures or delays the reaction of a child or class attempting a correct spelling. To what extent the variety of sentences, chosen extemporaneously by scores or hundreds of teachers during measurements of achievement in spelling, affects results, actually is not known. *E. g.*, In the report of the Butte survey (page 152) we are told that after the words were pronounced "and any explanations given by the teacher for their proper understanding by the children, the papers were collected," etc. (5) It is possible of course, that the failure to make uniform all sentences used illustratively with uniform tests in spelling is not a very serious defect in contemporary studies.

However, in the administration of a well-known list of words to about 25,000 children in 78 schools, the writer in advance wrote all illustrative sentences, and these were used by all examiners on the day of the test. The procedure was thus: The examiner pronounced the word, then spoke the illustrative sentence, whereupon the children wrote the word only. The list of words referred to is the Springfield, Illinois, test, devised by Leonard P. Ayres. (2) This list of words with the modifications referred to below was used as the basis of published reports on measurements of achievement in spelling by the school children of Butte, Montana; Oakland, California; Salt Lake City (4); Des Moines, Iowa, (1) etc.

More than one request has come to the writer for the list of uniform, illustrative sentences referred to above. The sentences are "standard" in the sense that we now have the results of their use as an aid during the administration of the Springfield List to large groups of children, classified as boys and girls, white and colored, and by grades. The data were obtained during the course of the

series of studies made by the writer in New Orleans during the spring of 1916. Although the results were given extensive local publicity it is believed to be worth while to present herewith for wider reading both (a) the illustrative sentences with the Springfield List, and also (b) the compiled results which afford the basis of tentative standardization. These results were tabulated by a group of paid and trained assistants who marked, and then checked all papers, and made the arithmetical calculations under close supervision. Schoolmen, it is hoped, may find the words and the sentences a convenient instrument for further measurements in spelling.

The tables of results are unusual in that the performances of two large groups of white and of colored children, taught in separate buildings, are contrasted. We do not know of the existence of a similar array of data obtained and analyzed with such care. Our impression, however, is that these data throw very little light upon the factors of race as affecting school achievements. The accompanying economic conditions, systems of instruction prevailing, physical equipment, etc., for the two races, white and negro, were markedly different. These are the condensed results classified by race:

TABLE I

Children Making Different Scores

The scores	100	90	80	70	60	50	40	30	20	10	0
The percentages of 17,642 white children	20	21	18	13	10	7	5	3	2	1	0.4
The percentages of 3,677 negro children	18	18	17	13	12	7	5	4	3	1	1

Here follow, first 60 words of the Ayres Springfield list with the illustrative sentences used by us, and then Table II containing the results by grade, sex, and race:

SPELLING SCALE USED IN SPRINGFIELD AND IN NEW ORLEANS WITH UNIFORM ILLUSTRATIVE SENTENCES (Ayres-Hill)

Grade 3.

FILL	Let us fill the pail with sand.
POINT	I can point to the North Star.
STATE	What state do you live in?
READY	We were ready just in time.
ALMOST	They almost missed their car.
HIGH	The kite flies high in the wind.
EVENT	We read of a very strange event.
DONE	The work was very well done.
PASS	The train will pass through the tunnel at night.
TUESDAY	We shall go skating on Tuesday.

Grade 4.

Forty	There were forty boys in the race.
Rate	The boat goes at a rapid rate.
Children	Today the children will have great fun.
Prison	The prison is so dark and gloomy.
Title	The story has a funny title.
Getting	We are getting our boat ready to sail.
Need	Fred will need more string for his top.
Throw	You watch when I throw the ball.
Feel	Do you feel well, Tom?
Speak	Let us speak to the little boy.

Grade 5.

Several	We shall use several yards of ribbon.
Leaving	They will be leaving for home tomorrow.
Publish	Why not publish this story in the school papers?
O'Clock	Our boat will sail at six o'clock tonight.
Running	Trout love to swim in running water.
Known	The story of Joe's trip was not known that day.
Secure	We shall secure an Indian guide to lead us.
Wait	Time and tide wait for no man.
Manner	She speaks in such an odd manner.
Flight	They quickly put the enemy to flight.

Grade 6.

Decide	Let us decide it at once.
General	I am sure it will help us in general.
Manner	He did it in such an odd manner.
Too	We must not hurry too fast.
Automobile	We shall have an automobile ride today.
Victim	He fell a victim to the enemy.
Hospital	We sent flowers to the sick child at the hospital.
Neither	You must neither look nor listen till I call.
Toward	The sun is moving toward the horizon.
Businsss	He did not like that kind of business.

Grade 7.

District	They may move to another district.
Consideration	He gave the matter very little consideration.
Athletic	He has a fine athletic form.
Distinguish	He can easily distinguish red from blue.
Evidence	He gave good evidence of being honest.
Conference	We must attend the conference this evening.
Amendment	The amendment received a large vote.
Liquor	They voted against the sale of liquor.
Experience	We had a jolly experience on the trip.
Receive	We shall receive the message by wireless.

Grade 8.

Petrified	We found some petrified wood.
Tariff	The tariff bill has passed in the senate.
Emergency	In an emergency I might be able to do it.
Corporation	The business is run by a large corporation.
Convenience	The automobile is a great convenience.
Receipt	We are in receipt of a note from James.
Cordially	She treated her little friend very cordially.
Discussion	The question raised a long discussion.
Appreciate	We always appreciate polite conduct.
Decision	He made a very wise decision.

TABLE II

RESULTS OF USE OF SPRINGFIELD LIST WITH UNIFORM SENTENCES

White

	—Boys—		—Girls—		—Pupils—	
Grades	No.	Pct.	No.	Pct.	No.	Pct.
8B	325	69.3	613	72.6	938	71.5
8A	257	77.5	432	78.2	689	77.9
8	582	72.9	1,045	74.9	1,627	74.2
7B	502	74.4	727	78.4	1,229	76.8
7A	343	78.6	500	78.9	843	78.7
7	845	76.1	1,227	78.6	2,072	77.6
6B	718	74.0	900	78.2	1,618	76.4
6A	517	79.5	668	85.7	1,185	82.9
6	1,235	76.3	1,568	81.4	2,803	79.2
5B	940	74.9	1,005	77.8	1,945	76.4
5A	710	81.1	779	85.9	1,489	83.6
5	1,650	77.5	1,784	81.3	3,434	79.5
4B	1,114	69.8	1,072	76.1	2,186	73.0
4A	868	78.4	832	82.7	1,700	80.5
4	1,982	73.6	1,904	79.0	3,886	76.3
3B	1,057	56.6	1,043	62.0	2,100	59.3
3A	895	69.9	825	75.1	1,720	72.4
3	1,952	62.7	1,868	67.8	3,820	65.2
Total Pupils	8246	....	9396	....	17, 642	....
Ave. Percentages		72.4		77.1		74.9

Colored

	—Boys—		—Girls—		—Pupils—	
Grades	No.	Pct.	No.	Pct.	No.	Pct.
8B	42	61.7	71	65.1	113	63.8
8A	18	70.0	40	75.5	58	73.8
8	60	64.2	111	68.8	171	67.2
7B	52	58.7	88	64.9	140	62.6
7A	17	70.6	45	69.1	62	69.5
7	69	61.6	133	66.3	202	64.7
6B	84	74.4	165	77.3	249	76.3
6A	52	83.3	102	83.0	154	83.1
6	136	77.8	267	79.4	403	78.9
5B	154	75.1	245	80.4	399	78.3
5A	76	83.0	133	84.2	209	83.8
5	230	77.7	378	81.7	608	80.2
4B	215	70.0	346	77.4	561	74.6
4A	201	73.5	238	82.0	439	78.1
4	416	71.7	584	79.3	1,000	76.1
3B	296	54.9	420	60.5	716	58.2
3A	261	68.7	316	74.9	577	72.1
3	557	61.2	736	66.6	1,293	64.4
Total Pupils	1,468		2, 209		3, 677	
Average Percentages		68.5		74.2		71.9

It is known, of course, and we are so cautioned by Ayres, that any list of words to be used for measuring certain attainments in school will become unreliable when children become accustomed to spelling the particular words of the scale. Investigators who may use the Ayres List will note, also, that our comparisons of this Springfield list of 70 words, contrasted with the list of 70 words actually used in the Butte, (5) Salt Lake City, (4) and Des Moines (1) studies, reveal slight variations from the original Springfield List, a matter not noted in the reports of any one of the three investigations mentioned above. *e. g.*, (1) Compare substitution of *organization* for *petrified* in Butte, (5) Des Moines (1), and Salt Lake City (4) studies; also *appreciation* for *appreciate* in Des Moines study. Furthermore, nine words in the Springfield List of 70 words (2) are not contained in the final list of one thousand words published in chart form (3) by the Sage Foundation. (*i. e.*, title, hospital amendment, liquor, petrified, corporation, convenient, tariff, leaving.) We are unable to explain the first irregularity, but a letter from Acting Director Richardson of the Sage Foundation stated in explanation with regard to the second "that the 70 words which you used were contained in the original list of 1,000 common words selected by Dr. Ayres, the words which appear in the original list were not all included in the thousand which were finally selected."

Rigid comparisons of the results of the Springfield, Butte, Salt Lake City, Des Moines results in spelling with the results obtained from lists drawn from Ayres' published chart (3) of the "1000 words most commonly used in writing" are obviously questionable. The usual assumption in giving the Springfield List, whether in its original or in the modified forms, is that these particular words were on the average spelled correctly by 70 per cent. of the children tested in about 100 cities. The lessons indicated are that we can not assume the identity of various lists of words originating in the Springfield List with words drawn from the Ayres published scale of 1,000 words, and that there is considerable room for improvement in the direction of standardizing a procedure for the actual administration of standard spelling scales.

REFERENCES

(1) *Annual Report of the Des Moines Public Schools*, 1915, 115 pp. See pp. 4-5.
(2) AYRES, LEONARD P. *The Public Schools of Springfield, Illinois.* Bulletin E137. Division of Education. The Russell Sage Foundation, New York City, 1915. 152 pp. See pp. 72-73.
(3) AYRES, LEONARD P. *A Measuring Scale for Ability in Spelling.* The Russell Sage Foundation, New York City, 1915, 58 pp.

(4) E. P. CUBBERLY and others. *School Organization and Administration. A Concrete Study Based on the Salt Lake City School Survey.* N. Y. 1917, 346 pp. See pp. 132-140.

(5) G. D. STRAYER and others. *Some Problems in City School Administration. Based on Butte School Survey.* New York, 1916, 234 pp. See pp. 150-155.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive reseaiches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3 00 pei year of nine monthly issues, payable in advance Single issues are 50 cents each Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X MAY—JUNE NUMBER 5-6

EDITORIAL

THE DYNAMICS OF MOTIVATION

Whatever of good or of evil the doctrine of interest and its advocates may have brought into American education, it is chiefly responsible for the present vogue of the term motivation. We hear much of motivated study, of the motivation of each step of the recitation, of the necessity for motivation in school conduct, and of the importance of motivation in all phases of education. This emphasis on motivation has the advantage of forcing the attention of the teacher from the text book ,the course of study or the class lesson plans to the learning process itself and to the attitude of the pupil toward school tasks. A fatal error of the old disciplinary education was the neglect of this matter of attitude or rather the assumption that the learning attitude could be elicited at command. So long as education was regarded as the privilege of the select few the evils of this assumption were ignored, but with the advent of the conception of universal education, and especially with the effort to make universal education compulsory, the need for a more careful consideration of the mainsprings of learning became evident.

Much of the current discussion of motivation in education is either superficial or mystical. What constitutes a motive, how are motives related to each other, how can motives be developed and utilized in educational procedure are questions that cannot be readily answered. An adequate psychology of motivation is yet to be written. One of the most stimulating contributions to the subject is to be found in Woodworth's recent book *Dynamic Psychology*. All explanations of conduct must be based upon the sensory-motor arc and its modifications. The action of the sensory-motor arc, even in its most highly complicated forms, is strictly mechanistic, but just as all mechanisms of which we have any knowledge require some application of force to actuate them, so the neural mechanism controlling behavior must have some impetus to set it going. Thus we have the fundamental distinction between "mechanism" and "drive." In the simple reflex arc the "drive" is furnished by the sensory stimulation. In the more complicated arcs involved in reflective behavior "drives" reinforce or inhibit each other, and undergo modification through memory and habit formation. Every "mechanism" must be affected by some sort of "drive" to produce a reaction, but any mechanism may act as a "drive" to bring about a reaction of any other "mechanism." Thus many "drives" result only in preparatory reactions, and the final, consummatory reaction, which is manifested in overt behavior, is the end station of an elaborate complex of initiatory "mechanisms" with their appropriate "drives."

The conception of "drive" lends itself admirably to the discussion of the emotional aspects of consciousness and their significance for thought and action. It is emotional attitude that is of the greatest importance in learning, and that needs much more careful consideration at the hands of educational psychologists. Educators need to study the "drives" of individual pupils, not so much to cater to their whims, as to build up those emotional complexes, those likes, desires, aspirations, that will constitute a motive power for the effort required in school learning. Recent studies in retardation and elimination point to the very great importance of the pupil's attitude toward his studies for his success in school, and it needs only casual conversations with pupils to discover the intensity of their likes and dislikes for different school subjects. Part of this attitude doubtless depends upon native endowment, and is little subject to modification, but the possibility of combining and developing "drives" in pupils suggests the results that might be attained by fostering a zest for learning.

J. C. B.

NOTES AND NEWS

In January of this year the Alpha Army Tests were given to about 3475 students of the University of Illinois. The medians for the different classes ranged from 121 to 163 (highest possible score 212). The highest score made was 207. According to the army ratings 105-134 was called B (Superior), and 135-212, A (Very superior). The vast majority of the officers tested made A and B ratings. The greater number of college students also made A and B ratings, very few having a score below 100. The women consistently made lower median scores than the men. The committee in charge of the application of the tests was David S. Hill (Chairman), B. R. Buckingham, C. A. Ruckmich and C. E. Holley.

The Otis Group Tests of Intelligence have now been made available in neat booklet form by the World Book Company, Yonkers, N. Y. The publishers state that the first printing of the tests was exhausted within a few weeks, and a much larger printing was called for than had been anticipated. It is to be hoped that Dr. Otis may be able to standardize the scores for the upper elementary and high school grades during the coming winter.

In view of the widespread professional interest in the reports constituting the Gary School Survey the General Education Board, 61 Broadway, New York City, announces that any volume will be sent free of charge on application. The list of reports is as follows: *The Gary Schools. A General Account,* Abraham Flexner and Frank P. Bachman; *Costs,* Frank P. Bachman and Ralph Bowman; *Industrial Work,* Charles R. Richards; *Household Arts,* Eva W. White; *Physical Training and Play,* Lee R. Hammer; *Science Teaching,* Otis W. Caldwell; and *Measurement of Classroom Products,* Stuart A. Courtis. The Board also announces for immediate publication and free distribution on request *Public Education and Private Endowment,* and *Public Education in Delaware.*

For the first time in the history of Harvard University properly qualified candidates for the Master's degree may present programs consisting of summer school courses only. This will be of obvious advantage to teachers who desire to become candidates for the advanced degree without giving up their regular positions. The Harvard School of Education has just announced a gift of $500,000 from the General Education Board, on condition that a sum of $1,500,000 be raised from other sources for the endowment of a Graduate School of Education.

The April meeting of the New York Society for the Experimental Study of Education dealt with "Problems Connected with the Junior High School." The program was as follows: "Problems in the Organization of the Junior High School," Professor Thomas H. Briggs; "The Relation of the Junior High School to the Senior High School," Dr. Joseph K. Van Denburg; "Practical Problems

in the Work of the Junior High School," Principal Robert B. Brodie. The May meeting was devoted to reports of the chairmen of committees on the projects contemplated for investigation next year.

The New York Branch of the American Psychological Association held its last meeting for the school year on May 5. The following papers were presented. "Experimental Test of the Newer Philosophy of Education," Dr. William A. McCall; "Correlation between Aesthetic and Recognition Value," Dr. Annie Berliner; "The Agreement between Judgments of Letters of Recommendation, Photographs, and Handwriting, and Judgments of the Success Achieved in the Educational Field," Miss Laura M. Chassell.

The Carnegie Institute of Technology, Pittsburgh, announces a number of scholarships, fellowships and research assistantships in the division of personnel and psychology. The scholarships carry remission of tuition, the fellowships a stipend of $300 to $600, and the assistantships a renumeration of $1200 to $2000. Many opportunities for employment as research specialists, field agents, statisticians, personnel directors, educational directors, and supervisors of personnel research with salaries ranging as high as $4000 are announced by the Institute.

Dr. Bird T. Baldwin, director of the Iowa Child Welfare Research Station, announces research assistantships in child psychology and in the nutrition of the child paying from $400 to $1500.

Mr. S. D. Porteus, formerly of Melbourne, Australia, well known for his maze tests of intelligence, has been made director of the research department at the Vineland Training School for the Feeble-Minded.

Professor Guy Montrose Whipple, professor of applied psychology and director of educational research at the Carnegie Institute of Technology, has accepted a position as professor of experimental education and director of the bureau of mental and educational tests at the University of Michigan.

Dr. Clinton P. McCord, health director for the board of education, Albany, N. Y., instructor in educational hygiene in the Albany Medical College, and consulting psychiatrist at the Berkshire Industrial Farm at Canaan, N. Y., has been giving courses in hygiene and physical diagnosis at the Cornell University Summer School of Physical Education.

Dr. Elsie Murray, professor of psychology and philosophy in Wilson College, Chambersburg, Penna., has accepted a similar position at Sweet Briar College, Virginia.

Professor Stephen S. Colvin, of Brown University, will lecture next year in the school of education of Boston University, giving four courses in the field of educational psychology. Edwin M. Chamberlin, Ph. D. (Harvard), has been appointed assistant professor of education.

PUBLICATIONS RECEIVED

FRED H. ALLEN. *Report on the Mentality of the Delinquent Boy According to the Binet-Simon Tests.* Twelfth Biennial Report of the Preston School of Industry, Ione, California, 1916. Pp. 22-31.

This group of 382 boys showed 19.4 per cent, of normal intelligence, 28 per cent. backward, 17.2 per cent. borderline cases, and 35.4 per cent. feeble-minded. The organization of the school endeavors to group the boys according to their mental level, and thus to give each one such training as he can profit by.

FRANK W. BALLOU. *A Plan for the Promotion of Teachers from Merit Lists.* Bulletin No. xiv, of the Department of Educational Investigation and Measurement, Boston, Mass., 1918. Pp. 63.

After a survey of the qualities of merit in teaching presented by Ruediger and Strayer, Boyce, and others, a scheme for the rating of teachers in Boston was devised, and its application to local conditions described in detail. The paper will be of value to those who are wrestling with the problem of the rating of teachers.

FRANK W. BALLOU. *Organization and Administration of Intermediate Schools in Boston.* Bulletin No. xvii, of the Department of Educational Investigation and Measurement, Boston, Mass., 1918. Pp. 75.

Section I of this bulletin is devoted to a general discussion of the intermediate or junior high school, section II traces the development of intermediate schools in Boston, section III considers both the elementary and secondary schools of Boston with reference to the need for intermediate schools, and section IV presents a plan for the organization and administration of intermediate schools. While the principle of special advancement classes and of adjustment of the work to the individual pupil is formally approved, the detailed time schedule for the VII and VIII grades does not give great promise for the effectual realization of this ideal.

HARRIET M BARTHELMESS. *Determining the Achievement of Pupils in Letter Writing.* Bulletin No. xvi, of the Department of Educational Investigation and Measurement, Boston, Mass., 1918. Pp. 35.

This bulletin presents the results of a test in letter writing given to VI, VII and VIII grade pupils in ten elementary school districts of the City of Boston. Letters were received from 3,603 pupils, and showed a median number of errors per letter of 7.1 for the VI grade, 6.4 for the VII grade, and 5.9 for the VIII grade. A detailed analysis is made of these errors with reference to the heading, the body of the letter, and the address. The illustrative samples of errors in grammar, phraseology and sentence sense afford interesting evidence of the difficulty that teachers have in developing good English habits.

G. VERNON BENNETT. *The Junior High School.* Baltimore: Warwick and York, 1919. Pp. xi, 224. $1.25.

This timely book presents in simple, compact form what many school men want to know about the junior high school movement. The author disclaims encyclopedic inclusiveness at the start, and relies largely upon his own experience and observation. A discussion of the problems presented in the transition from elementary to secondary education is followed by a history of the junior high school

movement. Various objections that have been offered to the junior high school plan are taken up and answered one by one, and its reflex effect upon the work of the first six grades receives careful attention. Several proposed courses of study are examined with reference to the emphasis placed upon the different subjects, much space is devoted to the character of the teaching force, the methods of teaching, and the administration of the work, and the relation between the junior and the senior high school is carefully considered. A final chapter sums up the author's views under the caption of an ideal junior high school. An appendix contains typical junior high school courses of study, and there is a bibliography of thirteen pages, arranged in the rather awkward fashion of grouping the articles according to the publications in which they appeared.

Clara Beverly and S. A. Courtis. *English Composition, for use in the Detroit Public Schools.* Detroit: Board of Education, 1917. Pp. 158.

The purpose of this monograph is to help teachers to use a composition scale by presenting a large number of samples from grades III to VIII evaluated on the basis of the Hillegas Scale. There are pertinent suggestions to teachers on the use of a composition scale, five sets of ten compositions each representative of the work of the Detroit elementary schools, thirty-five specimens from the low sixth grade evaluated by four trained judges, a rating of teachers' marks on the basis of the Hillegas scale, set J from the Thorndike collection of compositions, and three or four other sets of samples. The abundance of material offered will assist teachers who complain that the range of samples in the Hillegas scale is too restricted.

J. A. Bexell. *First Lessons in Business.* Philadelphia: J. B. Lippincott, 1919. Pp. 174. Sixty-eight cents.

Pupils of the eighth and ninth grades will find this little book of distinct value to them in mastering the factors of personality that make for success in business, and in becoming familiar with business forms. Among the topics discussed are honesty, courtesy, industry, personality, initiative, loyalty, dependability, thrift, bookkeeping, household accounts, business terms and forms, the business office, filing documents, and organization and management of savings banks.

Jean Broadhurst and Clara L. Rhodes. *Verse for Patriots, to Encourage Good Citizenship.* Philadelphia: J. B. Lippincott, 1919. Pp. xi, 367.

This splendid volume of poetry offers an ideal aid to the teaching of patriotism in the schools. From the verse produced during the war have been drawn some of the finest examples,—"The Road to France," Daniel M. Henderson; "Salutation, Marion Couthouy Smith; "In Flanders," John McCrae; "Verdun," Berton Braley; "The Soldier," Rupert Brooke; "I have a Rendezvous with Death," Alan Seeger; and many more.

Chester Lee Carlisle. *The Causes of Dependency, Based on a Survey of Oneida County, New York.* New York State Board of Charities, Bureau of Analysis and Investigation, Eugenics and Social Welfare Bulletin No. xv, 1918. Pp. 465.

In this excellent study we find a history of the county, a description of the people, an account of industrial development, an examination of children in the public schools, and a survey of the public institutions of the county from the point of view of dependency. The author presents evidence to show that the causes of dependency are chiefly hereditary. The report presents an abundance of case studies.

VOLUME X SEPTEMBER, 1919 NUMBER 7

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

A GROUP SCALE OF INTELLIGENCE FOR USE IN THE FIRST THREE GRADES

Studies from the Psychological Laboratory of Indiana University

LUELLA WINIFRED PRESSEY[1]

1. SPECIAL PROBLEMS INVOLVED IN THE DEVELOPMENT OF GROUP TESTS FOR YOUNG CHILDREN

The study described in the present paper began as an effort to devise a brief group scale of general intelligence for use (1) as a means by which children might be tested upon entering school and (2) as an aid in grouping these children into grades or sections, according to ability. It seemed to the writer that data of unusual interest might be gathered if such measurements of the entering classes -- the "pupil material" with which the school is to deal — might be obtained. Comparison of different schools as to the quality of their "pupil material" should be particularly interesting. However, such a scale might prove of greater value as a means of measuring the accuracy with which, in the first grades, children are selected according to ability for "fast" and "slow" sections, or are placed at entrance in their proper grade. In both these ways, then, such a scale should prove of unique value.

However, little work has thus far been done in the development of group tests, of either achievement or endowment, for use with such young children. Very early in the investigation, it appeared that work with first grade children involved certain special problems of test organization and technique, in the study of which the writer became quite as much interested as in the more specific effort to

NOTE 1—This paper is one of a series of studies being made by the Department of Psychology, Indiana University, in the field of mental tests. A later paper will deal with a scale complimentary to the examination described in the following paper, and intended for use in grades 4-8 inclusive. This second "Cross-Out" scale has already received brief notice in the Notes of this Journal.

develop a particular groups of tests. The paper should, then, be considered quite as much a study in the general method and technique necessary for work with young children, as a presentation of a particular scale particular set of methods.

As the writer analyzed the situation, the difficulties involved in the formulation of group tests for use with first grade children, were five in number. None of these difficulties appear in such prominence, if at all, in the making of tests for older children. In the first place, there were two special problems of material. (1) Older children can read and write; but no reading or writing, even of letters or numbers, can be involved in tests for young children. (2) Older children can work for some time without marked fatigue; young children cannot. It is very important that the examination be short and that the clerical labor of taking the examination be cut down to a minimum. Added to these two difficulties are three others having to do with the presentation of the tests. (3) Older children can understand and follow uninteresting and formal directions, if only they are clear. This is not true of little children. All kindergarten and first grade work is based on the principle of "spontaneous interest and participation" on the part of the children. This must be constantly kept in mind when giving tests to them. They must take an active part in the directions. In fact, their part must be the major one, the examiner merely guiding them through the examination. (4) Older children do not need a carefully integrated scale; they can easily shift from one problem to another, from one method of indicating score to another, or from one form of directions to another. Small children simply will not shift readily from one type of problem or reaction to another; there is a loss of interest and a period of general confusion after each test, before they are successfully launched on the next -- if, in fact, they ever do get started. Only by the most compact organization can the childrens' interest and attention be sustained from one test to the next. (5) Finally, older children will pay attention and do reliable work even if they are not particularly interested in the examination; habits of work and school discipline will keep them at their task. Primary children refuse point blank to work unless they are interested. There is here, in fact, a fundamental problem of motivation, which the writer considers of the greatest importance. Primary children are not accustomed to rigid discipline. They have not yet learned to work and simply cannot be

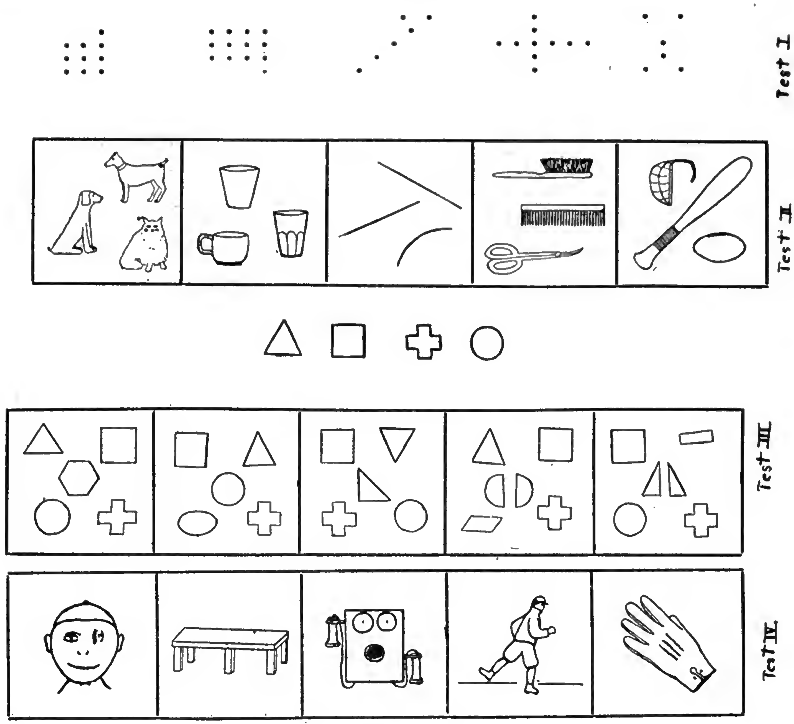
held to a task; they must be given work to which they will apply themselves of their own accord. This need of motivation from interest rather than compulsion – – so absolutely necessary in work with young children – – might, the writer feels, be considered in work with older children as well.

The writer's problem has been, then, to develop a scale which would meet these special problems, and would permit the measurement and comparison of entering classes in the schools and aid in individual diagnosis. The general line of attack may best be understood by a study of the tests finally incorporated into the examination, as given below.

2. Description of the Tests

The scale, as finally developed, consists of four tests, each of twenty-five items – – a total of one hundred items in all. In addition, there are five examples for each test. The first row of items for each test is reproduced in Plate I. In the first test, the children are asked to cross out the "extra" dot. In the second, they are told to cross out, in each square, the thing that is different from the other two things in the square. In the third, they are to cross out, in each square, the block that will be left over after all the other blocks have been fitted into the four forms at the top of the page. And, in the fourth, they are to cross out, in each picture, the part that is wrong. The examination appears on a four page folder with one test on each page, the pages being 11" by 8½."

(1) The studied avoidance of literacy in these tests is evident; two only of the hundred items involve any school acquirement and, in these two cases, the problem presented is not one wholly dependent upon direct teaching. In this respect, the writer feels that the examination is even more free from information gained from schooling than the Binet tests for the same ages. (2) Fatigue is kept negligible by calling for only one very simple response from the child. In each test he merely crosses out – – using any kind of a mark – – a certain object, part or dot. No fine eye-hand coordinations (so difficult for young children to make) are demanded, as any mark on approximately the right spot, will answer the purpose. By basing all four tests on this same "cross-out" principle the clerical labor of the child in taking the examination, is practically nil. Moreover, confusion from test to test is avoided. As a result of all these factors, it is possible to present the entire examination of 100 clear cut problems in less than twenty-five minutes; the



children actually work at each test only three minutes. (3) The children take an active part in the directions -- in fact, their answers constitute the *real* directions, as the majority of the children pay little attention to the final remarks of the examiner. The question-answer method is used. It might seem, in these days of absolute standardization of directions, that such directions would be loose and not capable of standardization. But, by using questions, the answers to which are very simple, the directions are kept practically standard. The examiner asks the same questions of every grade and from every grade, from the kindergarten thru the fourth, she receives the same, practically invariable, answers from the children. This method keeps the examiner in touch with the children, and has the further merit of automatically regulating the speed of giving the directions according to the time demanded by the class in understanding them. The directions for Test I are given verbatim below.

Look at the first group of dots at the top of the page. (Hold up copy and indicate general position). Can anyone see an extra dot -- one that is out of place?... Where is it?... Yes, above all the others. Now I want you to take your pencils and cross out that extra dot. Just draw a line through it. (Make sure that *all* the children get this first example correctly crossed out).

Now look at the next group of dots, just *beside* the first group. Where is the extra dot?... Yes, below all the others; everyone cross out the extra dot. Draw a line through it.

Now look at the next group, just beside the last one. Where is the extra dot?... Yes, above the line. Now all cross it out.

Now look at the next group. Where is the extra dot?... Yes, the last one on this side (gesture to the children's right). Cross it out.

Now look at the last group. Where is the extra dot?... Yes, the one just below the middle dot. Cross it out.

Now, everyone attention! Finish the other groups of dots on the page in the same way. In each group there is just *one* extra dot. Cross it out. (Time allowed pupils – – 3 minutes).

(4) The scale is carefully integrated. For each test, the five examples are in the same place on the page and are taken up in the same order; for each test, the directions develop these examples in the same way; for each test, the items are arranged the same way on the page. Above all, in each test, the child responds in the same way, – – he crosses something out. In fact, the most important element in the integration of the four tests is the use of this "cross-out" principle [2] in all the tests. (5) Most important of all is the development and sustaining of the childrens' interest throughout the examination. In the first place, the materials of the tests are pictures and blocks, in which children have a natural interest. The directions are given in such a way as to stimulate interest and rivalry among the children; the examples are interesting and the answers to the questions concerning them are so simple that every child can volunteer, thus having the pleasure of feeling he is contributing something to the success of the explanation. From the child's point of view, he and the examiner seem to be playing a game, in which he joins because he wishes to do so. The writer feels that there is a subtle but fundamental difference between an examination thus motivated by interest and the more formal type of examination usual in mental measurements.

3. Development of the Scale

Work on the scale was begun in the summer of 1918. A trial of various forms of tests was made with a mixed class of second and third grade children in the observation classes of the summer school of Indiana University. This class was composed of a few very bright children who were skipping a half grade and a few very dull children who needed extra training to pass the grade they had

[2] For a more complete discussion of the usefulness of this principle in test building, see Pressey, S. L. and L. W. *Cross-Out Tests with Suggestions as to a Group Scale of the Emotions.* J. of Appl. Psychol. June, 1919.

been in; it therefore presented an excellent group for experimental purposes. A tentative selection of tests was made from the results with these children.[3]

Experimentation during the fall was much hindered by the influenza epidemic, but some 200 children in country schools were tested and more systematic work was done in the primary grades of a small town. This experimentation served well to define the problem, make trial of methods and to lead to a more definite selection of the individual tests.

In the experiments thus far, the results had been obtained with pictures and forms drawn on the mimeoscope. The four tests that had been decided upon were now made into plates and blanks were printed for systematic study. In this form, the scale was given to all the children in the first, second and third grades of City A – – a city of about 12,000; 475 children were tested in all. The results were carefully worked over, the relation of the test findings to age-grade status, teachers estimates, Binet ratings and other criteria of ability studied, and the results of each test tabulated by items. On the basis of these findings a considerable number of new items were introduced (some entirely new, some taken from the earlier mimeographed sheets) to make the scale more evenly

[3] A total of some fifteen tests was tried at one time or another. Certain of these, though unsatisfactory for the writer's problem, might be of use to other workers and deserve brief mention. One (1) was a letter-pattern test in which the children were supposed to fill in the missing letters:—

L	M	C	B	A	B
L	M		A	B	A
L	M	C	B		B

The test was discarded because it demanded both knowledge of the alphabet and ability to write the letters. It should have possibilities, however, with older childen. Another test (2) involved discrimination of geometrical forms; rows of five forms, such as a square, a rectangle, a paralleogram, a rhomboid and a triangle, were presented. The children were asked to cross out, in each line, the form that was different from the other four. The test was rejected as being little more than a test of sensory discrimination. A third (3) test, of some interest was a directions test with pictures. The first item showed a face and the examiner directed that the children cross out the left eye; another picture showed a house and the directions were to cross out the window on the second floor that was not over a door or another window. The test was rejected because the children were so unused to disciplinary control that they would persist in looking around at their neighbors' papers between directions. A fourth test, (4) of some value for other purposes, was a substitution test. The key consisted of four geometrical figures, (a square, a cross, a triangle and a circle), with the silhouette of a lion in the first form, an elephant in the second, a kangaroo in the third and an eagle in the fourth. Below, on the page, were rows of the same forms, five in each row, with the animals again in the forms, but with one animal in each row in a different form from that in which he appeared in the key. The children were told to cross out, in each row, the animal that was in "the wrong cage." The children showed great interest in the test, but it was discarded because it was essentially a rate test and exact timing with young children is impossible.

graded in difficulty. The scale in this revised form was next given to all the children in the first three grades of City B; – – a total of 431 children. [4] Results from this trial were also tabulated by test and by item.

It was desired that the beginning of each test should be easy enough for any first grade child, that the end should be hard enough for use in the fourth grade if necessary, and that, between these two limits, the items should be arranged in order of difficulty, the steps between the successive items being approximately equal. After the analysis by item of the papers from City B was made, the items were graded on a probable error scale, using the second grade as a basis, and were then arranged in order of hardness. This treatment of the results showed that the steps were somewhat uneven and that there were some duplicate items (*i.e.* items of the same hardness). Other items, whose difficulty was already known from previous experimentation, were put in to help even the steps. The final product was a scale, the four tests of which were quite evenly graded in difficulty.

The results of this last revision were given to all the children in the first three grades of City C, – – with a total of 402 cases. The results of the survey with this final form are given below [5]

4. Results: First Norms and Validation

The age and grade distributions, with medians and percentiles, by total score, are given in Table I. The grade medians for the separate tests are given in Table II. These norms should be considered tentative as they are based on relatively few cases; but they will serve as a guide until further experimentation has been done and more cases tested.

---

[4] As indication of the ease and rapidity with which the examination can be given and scored, it might be said that all of these 431 children were examined by the writer and her husband in one day, between 8 a. m. and 3 p. m. and that all the blanks were scored by the writer and two students before 6 o'clock of the same day.

[5] Four items have since been changed because of difficulties in scoring. It is calculated that this slight change in items will raise the norms for the first grade about a point and between one and two points for second and third grades.

5-a. An intensive statistical study of the reliability and validity of the scale and separate tests is now being made. The data given below bearing upon the validity of the scale should be considered only a preliminary scale.

TABLE I

Primer Scale—Distribution of Scores

Children in grades 1–3 inclusive (402 cases) City C.

	I		II		III		Age				
Score	B	A	B	A	B	A	6	7	8	9	Score
95–99											95–99
90–94											90–94
85–89					1	2		1	1	1	85–89
80–84					4	5			4	2	80–84
75–79				2	5	16			9	8	75–79
70–74		2	3	4	5	17		7	7	12	70–74
65–59		2	8	16	12	9	1	15	16	10	65–69
60–64	2	8	4	13	13	14	5	14	12	15	60–64
55–59	2	4	5	8	5	4	6	7	7	5	55–59
50–54	7	9	12	13	4	4	6	22	13	5	50–54
45–49	5	17	8	6	2	5	13	14	9	5	45–59
40–44	5	11	4	2	4		10	5	6	2	40–44
35–39	5	5	7	2			8	7	1	2	35–39
30–34	5	5	2			2	8	4	3		30–34
25–29	3	6	1		2		6	2	3		25–29
20–24	2	2			1		2	2	1	1	20–24
15–19	7	3					6		3		15–19
10–14	6						5	1			10–14
5–9	6	3					6				5—9
0—4	11	1	1		1		9	1	1	1	0—4
Number	66	78	55	66	59	78	91	102	96	69	
	144		121		137						
Median	26.7	45.9	50.9	60.8	64.0	70.3	37.2	63.5	61.0		
25%	9.6	34.5	43.5	52.4	55.8	61.4	17.2	46.3	48.7		
75%	49.5	54.4	61.5	61.7	70.2	76.1	48.2	64.4	69.8		

Note: It should be noted that the curves for the IB and 6 year distributions are not normal because of the number of zero scores – – due to not understanding the directions. In calculating the medians for 7 and 8 years, the one seven year old and four eight year olds above the third grade,(and therefore not tested) have been counted in as being above the median. The other percentiles were calculated from the figures presented.

TABLE II.

Medians for each test.

Grade	No.	Test I	Test II	Test III	Test IV
1	144	10.3	10.5	9.2	8.9
2	121	13.3	15.5	15.8	14.7
3	137	15.7	17.0	20.3	16.1

In this last city all data possible were obtained bearing on the question of the validity of the scale. Light regarding the validity was gained from five sources: (1) correlation with teachers' estimates, (2) correlation with Binet ratings, (3) location on the total distribution of the scores made by children having a Binet I. Q. of 123 or more and by those having an I. Q. of 76 or less, (4) comparison of "slow" and "fast" sections, and (5) comparison of the entering classes in the different schools.

Seventeen correlations were obtained between the test scores and teachers rankings of pupils according to ability; these rankings were obtained from all teachers in City C, except from some few substitutes who did not know the children well. The average correlation was .47.

Three correlations were possible between the test scores and mental age on the Stanford Revision. For 64 unselected six year old children, the correlation is .64; for a group of 33 unselected seven year olds, it was .69; and for a similarly unselected group of 29 eight year olds, it was .79. [6] The product-moments method was used throughout. The writer feels that a higher correlation is hardly to be expected between a group test and an individual test - - at least in work with young children – – because the test situation is so different in the two methods. Bashful children do better when submerged in a group and lazy children do better when constantly prodded by the Binet worker, etc.

Among the 6, 7 and 8 year old children who were given Binet examinations in two of the three cities tested, 24 were found whose I. Q. was over 123. Twenty-two of these make group test scores in the highest ten per cent. for their age. The other two make very low scores that are obviously unreliable. Considering the difficulty of testing such young children at all, the scale seems reasonably reliable at the upper end of the distribution. There were only eleven children with an I. Q. of 76 or below; all of these made group scores in the lowest ten per cent. for their age. Since the two extremes of the distributions seem reliable, the scale would appear to be fairly satisfactory as an aid in individual diagnosis, as the cases to be further studied are usually located in the extremes.

Validation of a different sort, but of perhaps more direct implication, appears in the comparison of certain groups, such as sections and classes. In many of the classes tested, the teacher had already divided her class into "slow" and "fast" sections on the basis of her estimate of their ability. It is not to be expected that her judgment would be absolutely accurate; but the fact that a child has, throughout the year, been able to hold his place in a "fast" section when he might at any time have been transferred to a slower one, is surely of some significance. On the average, 76 per cent. of the children in the "fast" sections score above the median of the

[6] The ages were kept separate since it is obvious (in spite of frequent practice to the contrary) that combining several ages adds a spurious element to the correlation. Mental age was used instead of I. Q. since, for such a correlation, absolute score on both tests is necessary. No per cent. rating with the group rating was possible.

corresponding "slow" section and, in all classes but one, 100 per cent. of the children in the "fast" sections score above the twenty-five percentile for the corresponding "slow" sections.

Similar evidence appears when the children of a school whose pupils are known to be distinctly below average are compared with the children of another school that is known to draw from a superior neighborhood. Such evidence has been found in all three of the cities tested. In City C there are three ward schools: one of these is situated in the best residential district, another in a down-town district, and the third in the poorest part of the city. It had been felt for some years that the children attending the three schools were very different in their mental capacities – – that the "pupil material" of the three schools was very unlike. The results in Table III show the per cent. of children in the second and third schools who score above the median for the corresponding grade and age in the best school – – *i. e.* the one situated in the best residential district. If it can be assumed that the school officials were

Grade.	No.	Sch. 2.	No.	Sch. 3.	Age.	No.	Sch. 2.	No.	Sch. 3.
1	37	44%	48	40%	6 and 7	54	44%	57	42

right, the scale seems to have measured the differences already felt. It seems evident that the "pupil material" of the three schools is really different. In City B, there is a ward school that is attended almost exclusively by the children of university teachers and other professional men; another school in the same city is attended entirely by the children of factory and quarry workers. Only 24 per cent. of the children in the second school tested above the age medians of the children attending the first school; there was only one child in the second school who tested above the seventy-five percentile for children of his age in the first school. The second school has been considered for years as inferior to the first and, as was the case in City C, the scale has found and measured the difference. Similar facts can be presented for City A, in which 74 per cent of the children attending the "best" school score above their corresponding age medians for the children in the "worst" school, – – which in this case, draws from a neighborhood of the cheapest kind of quarry labor.

Such evidence as has been presented in the last two paragraphs is crude enough; but it is direct evidence and it is also an evidence of the usefulness of the scale in dealing with actual educational

problems. It shows that the scale is differentiating, to some extent at least, the ability that makes a pupil, or a group of pupils, better or less able than the average to do successful school work.

5. Discussion

The scale described in the previous pages is, as far as the writer knows, the first attempt at a group scale of intelligence for use with first grade children. It is, therefore, pioneer work. The writer feels that some discussion is needed concerning (1) certain special reasons why tests of intelligence for use with first grade children should be of unusual value and (2) certain special problems which appear in test work with young children and which need further study.

(1) Any school makes more or less adjustment to the mentality of the pupils attending it. That these adjustments may be as thorough as possible it would seem necessary to obtain some measure of the intelligence of this "pupil material." The only grade in which this measurement can be satisfactorily made is the first. In the upper grades some adjustment and considerable selection has already taken place; in fact, measurements of the upper grades may be little more than measurements of the extent of such adjustment and selection. [7] Comparison between schools on the basis of such results is impossible because of the different amounts of selection in the different schools. But that such differences are marked and important the writer has especially tried to show. Surely, measurement of the distribution of abilities obtained from the upper grades only gives a very partial and incomplete picture of the total situation in a school or school system. Scientific study of individual differences should be begun as soon as possible after the child's entrance.

(2) The writer feels that there are certain very special problems in building tests for young children. The most fundamental of these is the control of interest, but of nearly equal importance is the avoidance of confusion from test to test. Interest is controlled principally through the directions. There might be some doubt as to whether or not the question-answer method is the best way of doing it. This method keeps the examiner in touch with the class, keeps the class in touch with her and seems to work well. But other methods should certainly be tried. The writer is certain only of this, that in work with very young children a special problem is presented as regards directions. With older children it is enough

[7] Pressey, S. L. *A Comparison of Two Cities by Means of a Group Scale of Intelligence.* Educ. Ad. and Super. Vol. 5. pp. 53-62, 1919.

that the directions be clear; with young children there is the further fundamental requirement that there shall be control of interest. Any other methods used must recognize this fact.

Confusion from test to test has been avoided by close integration of the tests. Such unification greatly facilitates the work. It may be that, in thus simplifying the procedure, certain elements of difficulty which should appear in such an examination have been left out of the tests; the ability to shift from one test to the next may be one of the best indications of intelligence! But, at least, such unity is worth trying. It has seemed to the writer that it tended to a greater concentration of effort on the real problem of the tests, with less attention to irrelevant matters of test form.

6. Summary

The paper may be briefly summarized:

1. It is urged that a group scale of intelligence is necessary in the primary grades in order (a) to permit comparison of the "pupil material" entering different schools and (b) to aid in individual diagnosis.
2. It was found that, in building a scale for young children, certain special requirements must be fulfilled: (a) all elements of school training must be avoided, (b) fatigue must be reduced to a minimum, (c) the directions must be such that the children can take an active part in them, (d) the tests of the scale must be carefully integrated, and (e) the work must be motivated and maintained by interest rather than by disciplinary control.
3. A brief survey scale is presented which endeavors to meet these requirements. The scale consists of four tests, each of twenty-five items; the items are arranged in order of difficulty on a Probable Error scale.
4. Tentative norms are given, and evidence presented to show the validity of the scale as a measure of general ability.

A CLASSIFIED SCALE FOR MEASURING INTELLIGENCE

CARLETON W. WASHBURNE
San Francisco Sta.e Normal School

The rapidly increasing use of intelligence testing in schools necessitates a modification of existing tests. All tests now used measure intelligence more or less efficiently, but fail to diagnose in any consistent way the specific lacks or excellences of any individual intelligence.

School psychologists, to be worth their hire, must be able to say not merely, "This child is superior, this one inferior," but to say specifically, for instance, "This child is normal in most mental functions not involving immediate word memory, but he is deficient in that. Do not expect it in his work—build on the normal functions and see if you can train this deficient one."

All of us do, of course, make an effort at this sort of diagnosis, but existing scales for measuring intelligence are not adapted to our uses. For this reason I have undertaken to make a practical diagnostic scale which will not only measure the general intelligence but will also indicate in what particular functions a child is above or below normal.

The Problem and Plan of Attack

The problem in constructing the classified scale has been:

a. To make the test specifically diagnostic as well as general.

b. To make a classification which would be practical rather than technically psychological, and which would, therefore, be suggestive to teachers.

c. To have the results of the test such that they could be compared with the results of other widely used intelligence tests.

d. To use the results of the extensive investigations and revisions which have already been made by others.

e. To keep the test within the time and skill limits already set by successful intelligence measurement scales.

It is patent that the age-classification used so commonly is general and not diagnostic. The Binet scale was not planned for specific diagnosis. It set out deliberately to measure *general* intelligence. A scale should be devised which would cover the most significant

of the mental functions each in a thorough way. But as Wallin has pointed out, it is better not to complicate the field further by adding more revisions until a foundation or organization of psychologists undertakes the job and does it far more thoroughly than any one person or small group could possibly do it. Consequently I considered it better to use the best current revision of the Binet Scale as a basis for the classification, leaving the procedure, materials and method of scoring intact and simply rearranging the tests according to mental function rather than according to age. My classification is frankly a make-shift until such time as a new general revision can be made which will include adequate tests diagnostic of each function at different age levels.

Meanwhile, the revised Binet scale is fairly susceptible to a diagnostic reclassification. Yerkes and his co-workers, in their point scale, attempted one such reclassification. Their groupings, however, are distinctly technical. To the ordinary teacher and even to the trained psychologist they are not especially enlightening when it comes to giving practical suggestions as to the education of the pupil tested. Furthermore, the Yerkes scale came out just too soon to take advantage of the thorough work and revision carried on by Terman.

Terman's Stanford Revision of the Binet-Simon scale, in spite of several obvious faults, certainly is the best available general intelligence test for school use, and has an easily available, completely worked out procedure and scoring.

For this reason I have chosen the Stanford Revision as the basis for my diagnostic classification, and have so arranged my scale as to make use of Terman's procedure and scoring directions and to secure an I. Q. which is identical in each case with that which would be secured by the straight Stanford Revision.

The method of making the classification was necessarily rough. None of us knows exactly what mental functions are used in the passing of any one test. Different individuals doubtless vary somewhat among themselves in this particular, and quite possibly every mental act is conditioned by all mental functions. Nevertheless, different functions clearly predominate in different acts. My effort, therefore, has been to group together under the head of one function all tests the passing of which clearly involved the use of that function; to include a test in more than one group when more than one function was apparently of importance in passing it; and to make

the groupings not from the standpoint of technical psychology but rather from that of practical teaching problems.

In order to group the tests properly each test in the Stanford Revision of the Binet-Simon Scale was carefully examined. Terman's remarks on each test were considered and compared with the observations I had been making with this grouping in view. A temporary classification was made on this basis, then after about a hundred children had been tested by this preliminary classified scale, the results were studied and the classification thoroughly revised. The result was the following classified scale:

THE CLASSIFIED SCALE

(Note: Hyphonated numbers refer to Terman's age groupings, e. g., 3-1 means that the test so numbered is test no. 1 of year 3, of the Stanford Scale; 8–A1 is the first alternative test in the 8-year group, etc.)

No. *Title of test* *Grouping* *Cross reference*

I. NAMING THINGS

1. 3–1 Points to parts of body
2. 3–2 Names familiar objects
3. 3–3 Enumeration of objects in picture
4. (3–4) Gives sex .... See II, 1
5. 3–5 Gives last name
6. 5–2 Names colors
7. 6–1 Right and left .... Also II, 3
8. 6–5 Names three of four coins .... Also II, 4
9. 8–A1 Names six coins .... Also II, 9

II. COMMON KNOWLEDGE

1. 3–4 Gives sex .... Also I, 4
2. 5–A1 Gives age
3. (6–1) Right and left .... See I, 7
4. (6–5) Names three of four coins .... See I, 8
5. 6–A1 Morning or afternoon
6. (7–1) Number of fingers .... See XIII, 3
7. (7–4) Ties bowknot .... See XIX, 1
8. 7–A1 Names days of week
9. (8–A1) Names six coins .... See I, 9
10. (8–A2) Writes to dictation .... See XIX, 2
11. 9–1 Gives date
12. 9–A1 Names months of year
13 (10–4) Reading and report .... See XIX, 3
14. (12–2) Meaning of abstract words .... See XVI, 3
15. 14–3 Differences between president and king .... Also VII,8
16. (14–4) Problems of fact .... See IX, 7
17. (16–3) Differences in abstract words .... See XVI, 4
18. (18–4) Repeats thought of passage .... See XV, 17

III. IMMEDIATE AUDITORY MEMORY—DIGITS

1. 3–A1 Repeats three digits
2. 4–6 Repeats four digits
3. 7–3 Repeats five digits
4. 10–A1 Repeats six digits
5. 14–A1 Repeats seven digits
6. 18–3 Repeats eight digits

IV. IMMEDIATE AUDITORY MEMORY—SENTENCES

1. 3–6 Repeats 6-7 syllables
2. 4–A1 Repeats 12-13 syllables
3. (5–6) Three commissions............See XV, 4
4. 6–6 Repeats 16–18 syllables
5. 10–A2 Repeats 20-22 syllables
6. 16–A1 Repeats 28 syllables

V. VOCABULARY

1. 8–6 20 words
2. 10–1 30 words
3. 12–1 40 words
4. 14–1 50 words
5. 16–1 65 words
6. 18–1 75 words

VI. VISUAL MEMORY (PERCEPTION)

1. (6–2) Mutilated pictures............(See IX, 1)
2. 10–3 Copies designs from memory............Also XVIII, 6
3. (10–4) Reading and report............See XIX, 3
4. (12–4) Dissected sentences............See X,10
5. 14–6 Reversing hands of clock............Also IX, 8
6. 16-6 Code............Also X, 12
7. (18–2) Paper cutting............See IX, 10

VII. QUALITATIVE COMPARISON

1. 5–3 Aesthetic comparison
2. (5–5) Divided rectangle............See VIII, 3
3. (6–2) Mutilated pictures............See IX, 1
4. 7–5 Gives differences
5. (8–4) Gives similarities............See XVII, 1
6. (9–6) Rhymes............See X, 6
7. (12–8) Gives similarities, three things............See XVII, 4
8. (14–3) Gives differences between president and king............See II, 15
9. (16–3) Gives differences between abstract words............See XVI, 4

VIII. DISCRIMINATION OF FORM

1. 4–2 Discrimination of forms
2. (4–4) Copies square............See XVIII, 1
3. 5–5 Divided rectangle............Also VII, 2; X, 1; XI, 1; XVIII,2
4. (7–6) Copies diamond............See XVIII, 4

IX. IMAGINATION (PERCEPTION)

1. 6–2 Mutilated pictures............................Also VI, 1; VII, 3; X, 2
2. 7–2 Pictures, description
3. (8–1) Ball and field, inferior plan..................................See X, 3
4. (9–5) Three words in a sentence..................................See X, 5
5. (12–3) Ball and field, superior plan................................See X, 9
6. 12–7 Pictures, interpretation
7. 14–4 Problems of fact....................................Also,II,16; X, 11
8. (14–6) Reversing hands of clock................................See VI, 5
9. (16–4) Enclosed boxes....................................See XIV, 5
10. 18–2 Binet's paper cutting test....................................Also VI, 7
11. (18–6) Ingenuity test....................................See X, 13

X. INVENTIVENESS AND INGENUITY

1. (5–5) Divided rectangle....................................See VIII, 3
2. (6–2) Mutilated pictures....................................See IX, 1
3. 8–1 Ball and field, inferior plan..........................Also IX, 3; XV, 6
4. (9–2) Arranges weights....................................See XII, 3
5. 9–5 Three words in sentence..........................Also IX,4; XV, 8
6. 9–6 Rhymes....................................Also VII, 6
7. 10–6 Sixty words
8. 10–A3 Form board....................................Also XI, 2
9. 12–3 Ball and field, superior plan ....Also IX, 5; XV,12
10. 12–4 Dissected sentences..........................Also VI,4; XV, 13
11. (14–4) Problems of fact....................................See IX, 7
12. (16–6) Code....................................See VI, 6
13. 18–6 Ingenuity test..........................Also IX,11; XIV,6

XI. PATIENCE OR PERSISTENCE

1. (5–5) Divided rectangle....................................See VIII, 3
2. (10–A3) Form board.................................... See X, 8

XII. QUANTITATIVE COMPARISON

1. (4–1) Compares lines....................................See XV, 1
2. 5–1 Compares two weights....................................Also XV, 3
3. 9–2 Arranges five weights... ....Also X, 4

XIII. USING AND MANIPULATING NUMBERS

1. 4–3 Counts four pennies
2. 6–3 Counts thirteen pnniees
3. 7–1 Number of fingers....................................Also II, 6
4. 7–A2 Repeats three digits backward
5. 8–2 Counts backward, 20–1
6. 9–4 Repeats four digits backward
7. 12–6 Repeats five digits backward
8. 16–5 Repeats six digits backward
9. 18–5 Repeats seven digits backward

XIV. ARITHMETICAL REASONING (ALSO XIII)

1. 9–3 Makes change
2. 9–A2 Values of stamps
3. (14–2) Induction test......See XVII, 5
4. 14–5 Arithmetical reasoning
5. 16–4 Problem of enclosed boxes......Also IX,9; XVII,8
6. (18–6) Ingenuity test......See X, 13

XV. COMPREHENSION (ALSO IMAGINATION)

1. 4–1 Compares lines......Also XII, 1
2. 4–5 Comprehension, first degree
3. (5–1) Compares two weights......See XII, 2
4. 5–6 Three commissions......Also IV, 3
5. 6–4 Comprehension, second degree
6. (8–1) Ball and field, inferior plan......See X, 3
7. 8–3 Comprehension, third degree
8. (9-5) Three words in sentence......See X, 5
9. 10–2 Detects absurdities
10. (10–4) Reading and report......See XIX, 3
11. 10–5 Comprehension, fourth degree
12. (12–3) Ball and field, superior plan......See X, 9
13. (12–4) Dissected sentences......See X, 10
14. (12–5) Interpretation of fables (Score 4)......See XVII, 3
15. (16–2) Interpretation of fables (Score 8)......See XVII, 6
16. 16–A2 Comprehension of physical relations
17. 18–4 Repeats thought of passage heard......Also II,18

XVI. DEFINING (ASSOCIATION AND ANALYSIS)

1. 5–4 Definitions, use or better
2. 8–5 Definitions, superior to use
3. 12–2 Meaning of abstract words......Also II, 14; XVII, 2
4. 16–3 Differences between abstract words......Also II, 17; VII, 9; XVII, 7

XVII. ABSTRACTING COMMON ELEMENT–GENERALIZATION; INTERPRETATION

1. 8–4 Gives similarities, two things......Also VII, 5
2. (12–2) Meaning of abstract words......See XVI, 3
3. 12–5 Interpretation of fables, (Score 4)......Also XV,14
4. 12–8 Gives similarities, three things......Also VII, 7
5. 14–2 Induction test......Also XIV, 3
6. 16–2 Interpretation of fables, (Score 8)......Also XV, 15
7. (16–3) Differences between abstract words......See XVI, 4
8. (14–4) Problem of enclosed boxes......See XIV, 5

XVIII. MOTOR COORDINATION

1. 4–4 Copies square......Also VIII, 2
2. (5–5) Divided rectangle......See VIII, 3
3. (7–4) Ties bowknot......See XIX, 1
4. 7–6 Copies diamond......Also VIII, 4
5. (8–A2) Writes to dictation......See XIX, 2
6. (10–3) Copies designs from memory......See VI, 2

XIX. SPECIA TRAINING

1. 7–4 Ties bowknot..Also II, 7; XVIII, 3
2. 8–A1 Writes to dictation................................Also II, 10; XVIII,5
3. 10–4 Reading and report.........................Also II, 13; VI, 3; XV, 10

How Tests are Scored

As can readily be seen above, most of the tests fell under two or more heads. But in calculating the intelligence quotient the same test could not be scored twice, of course; neither could alternative tests be scored except where a regular test was omitted. To provide for this a simple device was used—two columns for scoring were made, one a months-credit column in which to record the number of months to be credited in calculating I. Q., and the other the function-age column in which to note what age level the pupil reaches in each function. In the function-age column there is a blank for score opposite each test every time it appears; but in the months-credit column there is only a blank for scoring each test once and no blank for scoring alternative tests. Thus the blank for scoring reading and report (XIX, 3 above) would occur but once in the months-credit column (under Special Training, XIX) but would appear four times in the function-age column, i. e., each time the test appeared in the scale (under Common Knowledge, II; under Visual Memory, VI; and under Comprehension, XV as well as under Special Training, XIX). This same device makes it easy to tell at a glance whether each test has been given once, for an unfilled blank in the months-credit column will immediately indicate that the test opposite has not yet been given. Naturally we only give each test once, the additional scoring in the function age column being done after the test has been completed.

Perhaps a sample page from the record booklet will make this easier to visualize:

Sample Page from Record Booklet

IV. IMMEDIATE AUDITORY MEMORY (SENTENCES)

Mo. Cr. | Fnct-Age

1)...* ...3–6. Repeats 6-7 syllables (1 of 3)
 a. I have a little dog..........................
 b. The dog runs after the cat...................
 c. In summer the sun is hot....................

2) ...4–A1. Repeats 12-13 syllables (1 of 3 absolutely correct or 2 of 3 one error each)
 a. The boy's name is John. He is a very good boy.
 b. When the train passes you will hear the whistle blow.
 c. We are going to have a good time in the country.

3) ...(5–6). Three commissions. (See XV, 4)

4)... ...6–6 Repeats 16–18 syllables. Instructions like 2) above.
 a. We are having a fine time. We found a little mouse in the trap.
 b. Walter had a fine time on his vacation. He went fishing every day.
 c. We will go out for a long walk. Please give me my pretty straw hat.

5) ...10–A2 Repeats 20–22 syllables. Instructions as above.
 a. The apple tree makes a cool pleasant shade on the ground where the children are playing.
 b. It is nearly half-past one o'clock; the house is very quiet and the cat has gone to sleep.
 c. In summer the days are very warm and fine; in winter it snows and I am cold.

6) ...16–A1. Repeats 28 syllables. (1 of 2 absolutely correct. (5 months)
 a. Walter likes very much to go on visits to his grandmother, because she always tells him many funny stories.
 b. Yesterday I saw a pretty little dog in the street. It had curly brown hair, short legs, and a long tail.

V. VOCABULARY

1)...* ... 8–6. 20 words...
2)...* ...10–1. 30 words...
3)...*† ...12–1. 40 words...(3 or 4 months)
4)...*† ...14–1. 50 words...(4 or 6 months)
5)...*† ...16–1. 65 words...(5 or 7.5 months)
6)...*† ...18–1. 75 words...(6 or 9 months)

VI. VISUAL MEMORY (PERCEPTION)

1)... ...(6–2) Mutilated pictures (See IX, 1)
2)... ...10–3. Copies designs from memory. (1 correct, other half correct.) Expose 10 seconds. (Also XVIII, 6) a...b....
3)... ...(10–4) Reading and report (See XIX, 3)
4)... ...(12-4) Dissected sentences (See X, 10)
5)...† ...14–6. Reversing hands of clock (2–3. Error not over 3 or 4 min.) (Also IX, 8) (*4 months*) 6.22... Time req.... 8.10... Time req.... 2.46... Time req....
6)...† ...16–6 Code. Writes "Come quickly," (2 errors, six minutes. Omission of dot counts half error. Illustrate with "war" and "spy") (Also X,12) (*5 months*) Errors...C. O. M. E. Q. U. I. C. K. L. Y. Time...
7) (18-2) ...Paper cutting (See IX, 10)

In the left-hand (months-credit) column of the booklet the number of months to be credited for each test is recorded whenever there is a blank line in this column opposite the test passed. The number of months to each test is the number assigned by Terman in his Stanford Revision—*i. e.*, two months for each test below the twelve-year level, three months for tests of the twelve-year level, four months for the fourteen-year level, etc. For the convenience of the examiner a mark † is placed by the months-credit blank opposite each test that is to receive more than two months credit, and at the end of each such test the number of months to be credited is shown in parenthesis, thus: (*4 months*) as in test VI, 5 above, or (*3 or 4 months*), as in test V, 3 above. In cases like the latter, where two numbers are given in the parenthesis, it is ordinarily the smaller of the two (*e. g.*, *3*) that is credited; but in case the examiner, to save time, is giving only those tests which are marked with an asterisk (*) he gives the larger of these two numbers (*e. g.*, 4) instead. In this last event—*i. e.*, when only the asterisk tests are give—all tests below the twelve-year level receive three months credit each instead of two.

In the right-hand (function-age) column, each test given is recorded either plus or minus according to whether the pupil passes it or fails to do so. The hyphenated numbers immediately preceding each test (*e. g.*, *3*-6), as above noted, indicate the age level and Terman's test number—thus 3-6 means test No. 6 (Stanford Scale) of the 3-year group. So the first of the two hyphenated numbers always shows the approximate age level (*e. g.*, three years) at which most children are able to pass the test opposite. The highest age level reached by a child in any function, therefore, indicates roughly his mental age in regard to the particular function under which the test is listed. This we call his "function-age" for that function.

Let us suppose, for instance, that we are testing a nine-year-old child. We come to the page in the record booklet just reproduced above. There is no test in Immediate Auditory Memory, Sentences, for a nine-year child. We try the ten-year test (no. 5) therefore. (For getting function-age the alternative tests, like this one, are often helpful, even though they play no part in determining I. Q.) He fails on it. We place a minus sign on the line in the function-age column opposite and pass back to the six-year test (no. 4). He passes it, but just barely. We note this and give him plus in the

function-age column. In the months-credit column we credit him with two months for this test (no. 4) and also with two months for test no. 1 of the same group—for his passing the six-year test in this function is clear evidence that he could pass the three-year test in the same function. It is not so clear that he can pass the five-year test (no. 3) Three Commissions) however, since he just barely passed the six-year test, and expecially as no. 3 involves other functions. Consequently we would either give no. 3 now or remember to do so when we came to it later on—the note "See XV,2" shows that we will find the complete test as no. 2 under group XV of the scale.

Passing down to the next group we give the vocabulary test.* Suppose the child we are testing makes a score of 45 words. This number would be recorded, he would be marked plus in the function-age column opposite the twelve-year test, no. 3, and would be given three months credit for this test in the months-credit column and also two months each for nos. 1 and 2. For here again, the passing of the higher test in the function is evidence of the ability to pass the lower ones, this time rigorously so. The score for the child for this grouping is therefore seven months credit, and his function-age is thirteen years (for his score of 45 words is midway between that demanded for the twelve-year level and that demanded by the fourteen-year level.

Next, we pass on down to group VI, Visual Memory. Here as in Group IV there is a big gap between the six-and ten-year levels. Following our former policy, we try him first on the test nearest his physical age of nine years. This is no. 2, Copying designs from memory. Suppose he passes it. We credit him in both columns and go on down the scale. The next test is Reading and report, but as the hyphenated numbers are in parenthesis and the note after the test says "see XIX, 3" we leave test no. 3 until we come to group XIX. Similarly we skip, for the time being, no. 4, Dissected sentences, and try no. 5, Reversing the hands of the clock. Suppose he passes this, we record it in both columns and go on to no. 6, the Code test, on which he fails. He now has credit for six months

*The vocabulary tests could rather easily be grouped under common knowledge, and might even be related to auditory and visual memory. But since they form a well developed series in themselves, their significance is, I think, easier to grasp by themselves than mixed in with other tests.

(two for No. 2 and four for No. 5) in the months-credit column, and a function age of fourteen. Then let us suppose that when we come to group X (not shown on the sample page), he fails on the Dissected sentence test of twelve-year level. In scoring him afterwards this will make a minus in the function-age column opposite test 4 of group VI (Visual Memory) in contradiction to the plus in test 5 of the same group—in other words, although test 5 indicates that he has a function age of fourteen years in visual memory, test 4 indicates that it is not even twelve. Of course this is probably due to the fact that both tests are conditioned by other and different functions, or it may be due to an incorrect grouping. Anyway, we score his function-age as 14—in visual memory, the minus sign signifying that while he has passed a test of the fourteen-year level in this function, he nevertheless has failed a test of lower age level under the same function.

The scoring, then, of the child we have imagined testing, would be, so far as this sample page is concerned, as follows:

Group	*Function-age*	*months-credit*
IV Sentences	6	4
V Vocabulary	13	7
VI Visual memory	14—	6

At the end of the entire test all months' credits and function-ages are tabulated in this manner in blanks provided on the front cover (see sample below). The months' credits are added together (with 24 months extra for the first two years for which there are no tests), and this sum is the mental age of the one tested. It is, of course, identical with that obtained in the regular Stanford scale. I. Q. is calculated as usual by dividing this mental age by the child's physical age. The summary on the cover shows at a glance the mental age and intelligence quotient of the child, then, just as in the Terman booklet. But in addition to this it shows an analysis of the I. Q., giving a practical diagnosis of the subject's strong and weak points, indicating in what functions he is normal, in what ones deficient, and in what ones superior.

The Significance of the Scoring

It should be recognized clearly, however, that to say a child is six years old in immediate auditory memory of sentences and fourteen in visual memory is only an approximation to the truth, an

indication rather than a complete diagnosis. But it is an indication with considerable practical value, and taken into consideration with other indications may shed valuable light on the pupil's educational needs. Let me take for example a real case:

H. B. was a girl twelve years old doing fifth grade work poorly. At the request of her teachers I gave her an intelligence test. It yielded an I. Q. of 91—too near normal to justify nearly two years retardation, especially in a school where an individual system of instruction is in use. The I. Q. therefore gave no clew as to the difficulty. The classified scale, however, showed her function-ages to be as follows

SUMMARY OF RECORD OF H. B.

(From cover of record booklet) Age, 12–1; Mental age, 11.0; I. Q., 91.

No.	*Test group*	*Fnc. Age*	*Mo. cr.*
I.	Naming things	+	14
II.	Common knowledge	9	4
III.	Repeating digits	+	4
IV.	Repeating sentences	+	4
V.	Vocabulary	8	2
VI.	Visual memory	9?	0
VII.	Qualitative comparison	—	4
VIII.	Discrimination of form	+	4
IX.	Imagination	12	7
X.	Inventiveness	12	11
XI.	Patience	+	
XII.	Quantitative comparison	+	6
XIII.	Manipulating numbers	+	10
XIV.	Arithmetical reasoning	14—	2
XV.	Comprehension	12	12
XVI.	Defining	—	4
XVII.	Extracting common element	16—	14
XVIII.	Motor co-ordination	+	4
XIX.	Special training	—	2
	First two years		24
	TOTAL MONTHS		132

Mental age 11 yrs. 0 months

Note: The plus sign is used wherever the gap in tests is such as to fail to give even an approximate function-age, but where no deficiency in this function is evident. The minus sign, not preceded by a number, indicates that while the gap in tests does not permit of an approximate function-age, there is an evident deficiency in this function.

Notes on H. B.: The functions in which she shows deficiency are common knowledge, vocabulary, defining, and special training (reading), all of which point strongly to poor environment, and visual memory and qualitative comparison where the lack of tests near her age make accurate scoring difficult—in each of these last two it was the failure in a single test which caused the apparent subnormality.

As just noted, H. B.'s lacks almost all pointed directly to poor home environment, while those functions which were largely innate, such as reasoning, extracting common element (induction), etc., were either normal or above normal. A further enquiry brought

out the fact that her father is an illiterate teamster, and that the girl's deficiency in common knowledge, vocabulary, etc., had led to marked diffidence and lack of self-confidence. A discussion of her case with her teachers gave them an immediate perception of the need for supplying the background of common knowledge, etc., which her home had failed to supply, and of bringing out her strong points in such a way as to restore her self-confidence.

The scoring in the function-age column admits of considerable latitude and discrimination, since it in no wise affects the I. Q. For example, in giving a bright 7-year old child the ball and field test, I noticed that as the child drew his path he made remarks somewhat as follows: "I'd go over here by the fence and look in the high grass, then I'd go over here and look in the bushes, then I'd look out in the middle," etc. The path was without plan and the test had to be scored minus in the months-credit column. But in the function-age column under imagination and under comprehension I gave the child full credit. Under ingenuity, however, I scored him minus in the function-age column as well as in the months-credit column. (The ball and field test is classed under all three functions, i. e., imagination, comprehension, and ingenuity). The difference between a complete failure and one which is only partial becomes, in this way, apparent on the score sheet, yet the rigorous determination of I. Q. is in no way vitiated.

The Results Achieved

For my own use, and that of my assistant, I have had mimeographed enough booklets to give the classified scale a satisfactory trial. A considerable number of children have now been tested by it, over a period of about six months covering a range from four years to fifteen (and a few adults) and in I. Q. from 73 to 165. The results of the tests have been very satisfactory—they have immediately suggested the type of work needed by the child and have helped materially in recommendations to teachers. The testing time, including the tabulation of results on the cover, ranges from twenty minutes for normal five-and six-year olds to an hour or more for very superior children of twelve and over. These times always include the full test, as I never use the short-cut test, indicated by asterisks, in my work. The test is as easily given as the regular Stanford Revision on which it is based. The scoring is only a little

more difficult and is very much more significant. The I. Q. yielded is of course identical with that yielded by the Stanford Revision—with this exception: The testing by the classified scale is likely to be more thorough and the I. Q. more accurate. For when there is a considerable gap in the Stanford Scale between two tests of the same function, as in the case of immediate auditory memory, for instance, a marked efficiency or deficiency in this function may never come to light. Such a child as we have hypothecated above, would perhaps pass all seven and eight year tests and therefore never be tested in the 6-year repetition of sentences in which he made such a poor showing. On the other hand, he might fail in all the twelve year tests and therefore never be tested in the 14-year test of visual memory, his marked superiority in this regard therefore never being indicated. In the classified scale the tendency is, so far as the tests will indicate this, to find the approximate age level of the child in each function.

Let me state once more that I frankly recognize that the grouping of the tests is only a rough approximation. With further testing it will doubtless have to be modified in various particulars. But I feel that there can be no question as to the advantages of testing intelligence by a classified scale rather than by the age-group scale heretofore used.

Conclusion

To sum up: (a) By classifying the tests of the Stanford Revision of the Binet-Simon scale according to the predominant function or functions used in passing them, it has been possible to make the scale specifically diagnostic as well as indicative of the general intelligence. (b) By making the groupings untechnical and practical the scale has been made suggestive to teachers. (c) By leaving the procedure and scoring of the Stanford Revision unchanged, the I. Q.'s yielded by the classified scale tests are strictly comparable with those obtained by the Stanford Revision, and (d) the results of Terman's extensive investigations have been used in their entirety. Finally (e) the test has proved itself to be within the same time and skill limits as those of the Stanford Revision.

The classified scale, therefore, solves in a practical way one important problem for the school psychologist—it shows him which particular mental functions are above or below normal and thereby gives one indication for the proper direction of educational effort.

A PRELIMINARY REPORT OF SOME GROUP TESTS OF GENERAL INTELLIGENCE

FRANCES LOWELL

University of Minnesota

In October, 1917, the writer began a survey[1] of all the schools in one of the representative counties in Minnesota, for the purpose of locating all cases of subnormality. Such a survey necessitates, among other things, visiting every school in the county, selecting all children of questionable mentality, and testing them, individually, with the Binet-Simon tests.

The selection of retarded and subnormal children above the third grade is relatively easy, for teachers' estimates, previous school records, and repetition of grades without the excuse of a long continued absence, taken together form a fairly accurate basis. However, such methods are practically impossible in and below the third grade. This is due to several factors: first, to the inaccuracy of the school records for primary grades; second, to the frequent change of teachers, especially in the rural schools; third, to the fact that many children have been in school too short a time to become sufficiently retarded pedagogically to suggest a diagnosis of mental deficiency; and fourth, to the inability of many children, on entering school, to do themselves justice in the grade because of the strangeness of the new environment. Other difficulties, common to all grades, are also encountered in such a survey. Teachers frequently persuade themselves that since a child is sweet, docile and attractive, he must necessarily be bright; or, because he is doing good work in his grade, he is undoubtedly normal, regardless of the fact that he is two or three years older than his classmates. In one rural school, the teacher, on being asked about the work of twin girls in the third grade, said with great assurance, "Oh, those girls are very bright. They do good third grade work, and they are such nice children." The twins in question had a chronological age of twelve years, a fact entirely overlooked by the teacher in her estimate of their ability. Mentally, they were found to be 7 5-8 and 8 3-8 years, respectively, when given the Binet tests. Hence, in order to have an accurate survey, the need for an impersonal method of selecting children of doubtful mentality was imperative.

To supply this need, the writer has devised a series of group tests, to be used in the first, second and third grades, in determining which children should be given the individual Binet tests. Some of these group tests are adaptations of the original Binet-Simon tests; some have been taken from Dr. Kuhlmann's 1917 Revision; others from Dr. Terman's Revision; still others were suggested by tests standardized by the Bureau of Analysis and Investigation in New York, and the rest are original.

The Selection of Group Tests. The requirements of the group test are many, for not only must it possess the characteristics necessary to the individual test, but it must also satisfy various other demands. Simplicity is an important criterion in the selection of the group test; simplicity of material; of directions; of response; and of scoring.

Since a large number of children must be tested at one time, it is necessary that tests be selected in which the material used can be easily carried and quickly distributed. In order to facilitate handling, and to avoid confusion, the writer has arranged the material for the group tests in booklets. To each child, who has been in school less than one year, *i. e.*, who is in the first grade, and hence can read and write but little, a booklet, 7½ in. x 7½ in., is given, which contains material and blanks necessary for the fifteen tests designated as five, six and seven year tests. Children in the second grade receive similar booklets designed for six, seven and eight year tests; while third grade pupils get booklets for the seven, eight and nine year tests. Thus, all the responses of each child are kept together, and the confusion incurred by distributing and collecting loose sheets of paper every few minutes is avoided. All pages of the booklet are numbered, and definite spaces are alloted to each test. All possible elimination of disturbing elements is important, for in the primary grades the attention of the children is easily distracted.

Not only must the material for the tests be simple, but the directions to be followed must be clear and brief. Herein lies the greatest difficulty in selecting tests for a group. Frequently one feels confident that the directions for a certain test are perfectly clear and that they could not possibly be misunderstood, and yet, when the test is given to a group of children, he finds that the meaning is entirely lost. Good English must frequently be sacrificed, for the child in the primary grades is surprisingly limited in vocabulary.

Originally, in giving the directions for one of the six years tests, the writer said, "Make a cross in the LARGEST square." The result was puzzling, for the children crossed the smallest square as often as they did the largest. The test apparently was a failure for that age group. However, before discarding it, the writer decided to experiment a little to see if the difficulty could be discovered. Instead of having the children cross the square, they were asked to point to the largest square, whereupon one little girl tearfully informed the writer that she "didn't know what that meant." That solved the problem. From then on the children were instructed to "Make a cross in the BIGGEST square," and the success of the test was assured. Brevity is equally essential in giving directions to children, for their habits of thinking, of concentrating the attention, are not yet formed, and to listen, comprehendingly, to long directions is impossible for them. Novelty and change are so obviously important in maintaining the interest and stimulating the child to his best effort, that further discussion of the value of brevity is unnecessary.

In the selection of group tests, only those should be chosen which permit of but one correct response, and that response must have but one possible interpretation. It is true that some of the tests consist of several trials of the same sort; thus, three different series of five digits each, are read to the children, to be reproduced, but the child passes the test if he succeeds in reproducing any one of the series correctly. This is a necessary procedure, because it is almost impossible, in a group, to have perfect silence while the directions are being given; and so a single series might not be heard by all. Then, too, many children have poor auditory imagery, which fact makes it difficult for them to reproduce material presented orally. However, in all of the tests, a response is either entirely correct or it is entirely wrong. This "All or none" method of scoring is used in the writer's group tests, for the purpose of the series is not to discover how much better one child can perform a given task than another, but to see which children fail to perform it at all, *i. e.*, to sort out the subnormal children who need to be given the individual tests before a correct diagnosis can be made regarding their mentality. That the responses must permit of no variation in interpretation, due to the examiner's judgment, is undoubtedly as essential in group tests as in the individual tests.

Altogether, twenty-five tests have been selected which seem to fulfill the requirements mentioned above. These have been divided into groups of five, designated for convenience, as five, six, seven, eight and nine year tests. These have been given to 566 school children distributed according to age as shown in Table I.

Table I.

Age		*No. Cases.*
6	(i. e., 5 yrs. 6 mos. --- 6 yrs. 5 mos. incl.)	74
7	(i. e., 6 yrs. 6 mos. --- 7 yrs. 5 mos. incl.)	132
8	(i. e., 7 yrs. 6 mos. --- 8 yrs. 5 mos. incl.)	155
9	(i. e., 8 yrs. 6 mos. --- 9 yrs. 5 mos. incl.)	115
10	(i. e., 9 yrs. 6 mos. --- 10 yrs. 5 mos. incl.)	90
	Total	566

GROUP TESTS

Tests for Year V.

V, I. Formboard

Materials.

This is an adaptation of Goddard's [1] formboard. From heavy white cardboard the forms shown in black in Fig. I are cut. Black paper is pasted over the back of the openings to make them stand out more clearly, and then another sheet of white cardboard is pasted over all this to form the back. Ten little cards which exactly fit the ten forms, are placed in an envelope pasted on the back of the formcard.

Procedure.

The examiner holds up a booklet in front of the group of children to be tested, and says:

"Open these little books to this black and white card. In the envelope on the back you will find a number of little cards. Take these out of the envelope and put them in a pile on your desk. Now, there is one place in the black and white card into which each of these little cards will exactly fit, — like this. (Illustrate with square). See if you can find the place for each of the others."

Help may be given any child who has difficulty in finding the place for each card the *first* time. When all the children have the cards properly placed, have them remove them and again place them in a pile on the desk. Then proceed as follows:

"Now you know where each little card belongs. This time we are going to see who can put all the cards in their places *first*. Everyone wait until I say 'GO,' and then work as fast as you can until I say 'STOP.' Then put your hands in your laps AT ONCE, without touching a single card again. Is everybody ready? GO."

Allow 50 seconds before giving the signal to stop. Care must be taken to see that no child touches a card after he has been told to stop.

Scoring.

The test is passed if all ten cards are in their proper places at the end of 50 seconds. A plus or minus, indicating whether the child has passed or not, can be quickly marked on the blank page opposite the form card, by the examiner, before having the children replace the cards in the envelope.

[1]This survey was made under the direction of Dr. Fred Kuhlmann. Director of Research, Faribault, Minnesota.

[1]H. H. Goddard. *The Training School.* Vol. IX, No. 4, June 1912, pp. 49-54.

V, 2. Counting 4 Circles. (Lowell)

Materials.

A card, 6 in. x 18 in., on which are printed four large circles about 2 inches apart, (See Fig. VII), is necessary, and also two wooden blocks or other small objects. Space for the child to make the crosses is provided at the top of page 2 of the booklet.

Procedure.

"When I say, 'Make a cross,' this is what I want you to make. (Make X on the blackboard). If I should say, 'Make a cross for each of these blocks, (holding up two blocks), how many crosses would you make? Yes, two,—one for this block, and one for that, (XX), see?

"On this large card are some circles. When I hold the card so you can see them you are to count the circles TO YOURSELVES, and then make as many crosses up here at the top of your paper, (Indicate), as there are circles. Ready. Look." Expose the circles for 10 seconds.

Scoring.

The test is passed if four crosses are made.

V, 3. Copying Square. (Binet)[1]

Materials.

Use a large cardboard on which an eight inch square is printed in heavy lines.

Procedure.

Holding up the large square, and indicating the space in the booklet to be used, the examiner says:

"At this side of the big space, here, see how nicely you can make a square just like this one."

Hold the card so all can see it while they are drawing, and then when they have finished one, say:

"Now see if you can make a still better one at this other side of the big space."

Scoring.

The test is passed if one of the two squares drawn is as good as those in the score card used by Binet.

V, 4. Discriminating Colors. (Binet)

Since it is impossible to have children in a group test name colors, the following adaptation of Binet's color test was used. It necessarily makes a different test from the original, since the directions must be comprehended and followed exactly.

Materials.

One-half inch squares of the colors red, yellow, blue and green, are pasted on page 2 of the booklet in the order named.

Procedure.

"Look at the colors in the middle of this page. Listen carefully, and then do just as I ask you:

1. Make a cross *in* the *yellow* square, like this.
2. Draw a line *through* the *green* square, like this.
3. Now make a cross *above* the *red* square, like this.
4. Draw a line *under* the *blue* square, like this.

---

1 Annee psychologique (1911) vol. 17, pp. 145–201.

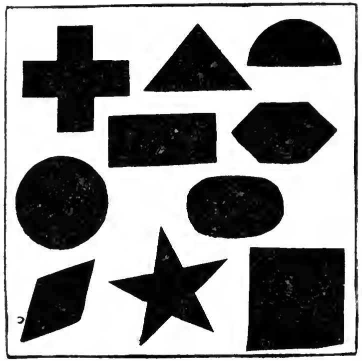


Fig. I

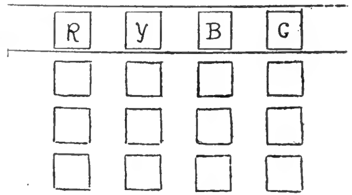


Fig. II

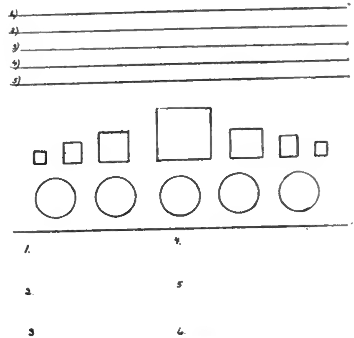


Fig. III

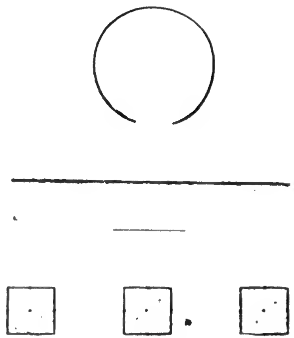


Fig. IV.

Fig. V.

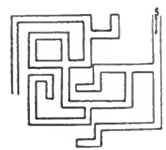


Fig. VI.

Fig. VII

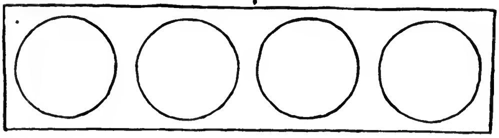


Fig. VIII

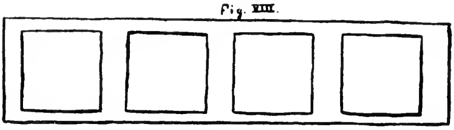


Fig. IX

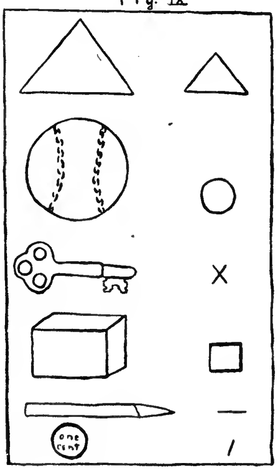
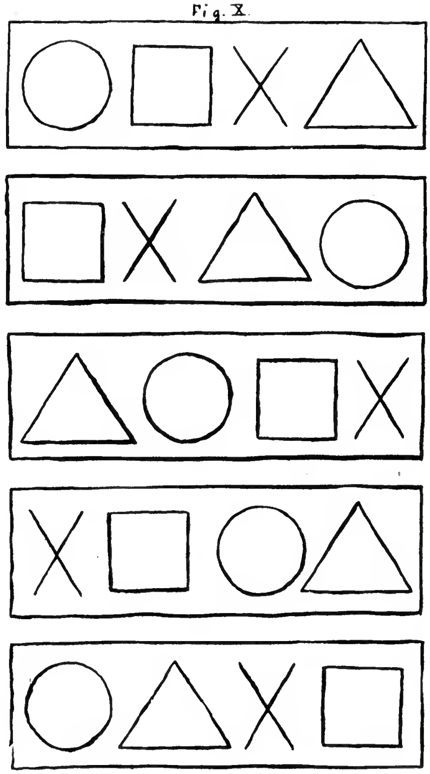


Fig. X



In order that the child shall understand what is wanted, draw a square on the board each time, and do what you tell the child to do. Give directions slowly, and repeat each.

Scoring.

The test is passed if all four colors are correctly marked.

V, 5. Irregular Tapping. (Pintner;[1] Kuhlmann.[2])

The present adaptation has been made from the form used by Dr. Kuhlmann in his 1917 revision of the Binet Tests.

Material.

Use a card, 6 in. x 18 in., on which are printed four squares, about two inches apart. See Fig. VIII. At the bottom of page 2 in the booklet, are three rows of squares similar to the one on the examiner's card. See Fig. II.

Procedure.

"At the bottom of the page you see three rows of squares. The first row looks like this one, doesn't it? (Point to card in hand). Now I am going to tap some of these squares, like this. (Tap the first two squares with a ruler). Watch me carefully, and when I've finished tapping, put a cross, like this, X, in just the squares that I tap. Ready. Watch."

Then tap the squares in Series A in the order named, at the rate of one square per second, and allow time between series for the child to cross the squares tapped. Always tell the children which row of squares to use for each series.

A) 1–2–4
B) 1–3–4
C) 2–3–4

Scoring.

The test is passed if one of the three rows is correctly marked.

Tests for Year VI.

VI, 1. Aesthetic Comparison. (Binet)

Material.

On each of the pages 3, 4, and 5 of the booklet, are two pictures used by Binet for the comparison.

Procedure.

"Look carefully at the two pictures on this page (3). Make a cross under the prettier one of the two."

Give the same instructions for each of the pages 4 and 5.

Scoring.

The test is passed if the prettier face is indicated in two of the three series.

---

1 Rudolph Pintner and Donald G. Paterson. *A Scale of Performance Tests,* 1917.

2 Fred Kuhlmann. *The Measurement of Mental Development,* A School Publication, —Faribault—1917.

VI, 2. Mutilated Pictures. (Binet)

Materials.

The four pictures used by Binet are printed on four successive pages in the booklet, 6, 7, 8 and 9.

Procedure.

"Look at this picture. (page 6). Part of the face is gone. Let us see what part it is. You see the eyes. Look, the eyes are there. You see the nose. Yes, the nose is there, and look, the chin is there. Now what part is gone? (Children name mouth). Yes, the mouth is gone. Now make a cross right where the mouth should be, to show that that is the part that is gone.

"Look at the next picture. Do not tell me this time what is gone, just make a cross to show me what part is missing.

"Make a cross to show what is gone from the next picture. (Page 8) "And now show what is gone in the last picture." (Page 9).

Scoring.

The test is passed if the last three pictures are correctly marked.

VI, 3. Counting Irregular Taps. (Kuhlmann).

Materials.

The examiner needs a wooden block or the blunt end of a pencil for tapping. Ruled blanks are provided the children on page 10 of the booklet. See Fig. III.

Procedure.

"Listen. I am going to tap on the table and see if you can count the numbe of taps. You count to yourselves, and give me the number when I ask you. Ready."

Tap 5 times, at the rate of one tap per second. See that the hand is screened while tapping, by a large cardboard.

"How many times did I tap? Five, that is right. Now, instead of telling me how many times I tap, I want you to make a cross for each tap, when I get all through. I just tapped five times, so how many crosses shall I make to show it? Yes, five, like this: XXXXX.

"Sometimes I will stop tapping and then begin again. Don't let that fool you. You count only those you hear, and when I ask how many, you will make one cross for each tap, here, on this first line. (Indicate line each time). Ready. Listen."

Tap the following series in order, at the rate of one tap per second, for each cross, and pausing a second for each dash:

1)	X X - X X	(4)
2)	X - XX - X X X	(6)
3)	X X X - X	(4)
4)	X X - X X X - X	(6)
5)	X - X X - X - X	(5)

Scoring.

The test is passed if 3 out of the five series are correctly counted as indicated by the correct number of crosses.

VI, 4. Two Simultaneous Commands.

This test was suggested by Binet's test in which three simultaneous commands were used.

Materials.

One row of circles and a row of squares of different sizes are provided for the children on page 10 of the booklet. See Fig. III.

Procedure.

"See these squares and circles. Listen carefully, and see if you can remember what I am going to ask you to do with them, when I am all through telling you.

"Make a cross in the BIGGEST square.
Then draw a line in the FIRST circle."

"I will tell you once more, and then see if you can do both things. Listen:

"Make a cross in the BIGGEST square.
Draw a line in the FIRST circle."

Scoring.

The test is passed if both directions are correctly followed.

VI, 5. Perception of Sound. (Lowell)

Material.

A bean-bag, a baseball, a penny, a pencil, a key, and a wooden block are used. Have a table which can be concealed from the children and from which the above named objects can be dropped. A card picturing each of the objects with the respective symbols to be drawn by the children, is tacked up in view of the children. See Fig. X. Space for six drawings is provided at the bottom of page 10 in the booklet. See Fig. III.

Procedure.

"What is this?"

Hold up each of the objects and have the children name it.

"Now look at this card, (Fig. X), and tell me what each of these pictures is." (Have them name them all).

"Now, I am going to drop each one of these on the floor where you cannot see it, and I want you to listen to the noise each makes, and see if you can tell which of these I drop.

"If I should drop the bean-bag first, you would draw, in this first space, here, this. (Point to symbol at the side of the picture of the bean-bag); if I dropped the key next, you would make this X in the second space, here; if I dropped the penny next you would make this line in the third space, etc.

"Now shut your eyes, and see if you can tell from the sound which one of these I drop, and then look at the card to see what to draw. Then draw that in the first space. Ready, listen."

Drop objects in the following order:

block
bean-bag
penny
key
ball
pencil

Scoring.

The test is passed if five out of the six objects are named in the correct order.

TESTS FOR YEAR VII.

VII, 1. Number of Toes.

Because it is impossible to prevent a large group of children from counting their fingers if one were to use the original Binet test, the number of toes on each foot and on both feet together, has been substituted.

Procedure.

"I am going to ask you some questions, and I want you to write the answers, here at A) – – –, at B) – – – and at C) – – –. Ready:

A) How many toes on your left foot?
B) How many toes on your right foot?
C) How many toes altogether on both feet?"

Scoring.

Since most children of seven years can make figures up to 10, further use of crosses in answering questions is not permitted. The test is passed if the answers to all three questions are correct.

VII, 2. Memory Span. (Binet).

Procedure.

"I am going to read you some numbers. Listen carefully, and see if you can write them when I get through, just the way I read them."

Read the digits at the rate of one digit per second, without rhythm.

"Listen:

A) 6–5–2–8–1
B) 4–9–3–7–5
C) 2–8–6–1-9

Scoring.

The test is passed if the digits in one series are correct, even though their order be changed.

VII, 3. Copy Diamond. (Binet).

Material.

Use a card on which a large diamond, 9 in. x 18 in., has been drawn in heavy lines.

Procedure.

"At the left side of this big space, see how nicely you can draw a diamond just like this one." Hold the card so that all can see it as they draw. After they have finished drawing the first diamond, say:

"Now draw another one at the right side of the big space just like this one."

Scoring.

The test is passed if one of the two drawings is as good as those on the score card used by Binet.

VII, 4. Tying Bowknot. (Terman)[1]

Material.

Two pieces of tape, each about seven inches in length, are fastened to a stiff card pasted at the bottom of page 11, in the booklet.

Procedure.

"You know what kind of a knot this is, don't you? (Show them bowknot already tied). Yes, it is a bowknot. I want you to tie the same kind of a knot with the pieces of tape you have, just as quickly as possible."

Scoring.

The test is passed if a bowknot is correctly tied.

VII, 5. Geometrical Figures. (Lowell)

Material.

Use five cards, on which are printed, in a different order, on each card, a circle, a square, a cross, and a triangle. See Fig. IX.

Procedure.

"I shall show you some cards for ten seconds each. On each card you will find these forms. (Draw on blackboard). But the order in each case will be different. On one card a square will come first, and on another card the circle will come first, etc. You will notice carefully the order of the forms, and when I take the card away, you will draw what you saw in just the order it was on the card. Draw as quickly as possible. Ready. Look at the first card."

Allow not more than one minute for reproducing a card. Be sure all children are giving attention when you hold up a card.

Scoring.

The test is passed if three of the five series are correctly drawn.

Tests for Group VIII.

VIII, 1. Ball and Field. (Terman).

Material.

The incomplete circle used by Terman to represent the baseball field, is printed in the booklet.

Procedure.

"Let us suppose that your baseball has been lost somewhere in this round field. You have no idea what part of the field it is in. All you know is that the ball is lost somewhere in this field. You are to mark out a path to show me how you would hunt for the ball so as to be sure not to miss it wherever it might be. If you draw just a line to the center from the gate, like this, — it would mean that you only looked that far and then stopped, — and the ball might be away over at the other side; you don't know where it is. Now begin at the gate, and with your pencil mark out a path to show how you would hunt for the ball so as to be sure not to miss it."

In giving an individual child this test, according to Dr. Terman's directions, the examiner watches what the child draws, and if he stops after drawing just a short path, the examiner asks him where he would go next if he hadn't yet found the ball. As such procedure is obviously impossible with a large group of children, the illustration is used in the directions.

1 L. M. Terman. *Measurement of Intelligence.* 1916—pp. 196–199.

Scoring.

The test is passed if the child shows a plan as good as those shown on Terman's score card for eight year old children.

VIII, 2. Writing from Dictation. (Binet)

Procedure.

"I want you to write something for me as nicely as you can on this long line. Write these words: 'See the little boy.' Be sure to write it all: 'See the little boy.'"

Scoring.

The test is passed if the sentence is written legibly enough to be easily read, and if no words are omitted. Unless a word is so incorrectly spelled as to make it unrecognizable, the error is not counted.

VIII, 3. Counting the value of Stamps. (Binet)

Material.

Use a cardboard on which have been pasted three one-cent stamps, in a row at the top, and three two-cent stamps in a row under these.

Procedure.

"Write the answer to this question on the short line, here." Hold the card before the children and say:

"How much will it cost to buy all these stamps? A green stamp is worth one cent, and a red stamp is worth two cents. Now, how much are they worth altogether?"

Scoring.

Test is passed if the answer, 9 cents, is given.

VIII, 4. Comprehension Test. (Kuhlmann).

Material.

Three one-inch squares are printed in the booklet.

Procedure.

Point to the parts as you mention them.

"See this first square. This is the *center* of the square. What is it? This is the *upper right corner* of the square. What is it? This is the *lower left corner*. What is it? This is the middle of the *left side of the square*. What is it?

"Now take your pencils, and draw a straight line from the center of the first (second and third) square to:

1) The upper left corner.
2) The middle of the upper side.
3) The lower right corner of the square."

Scoring.

The test is passed if two of the three lines are correctly drawn.

VIII, 5. Estimation of length of lines. (Kuhlmann).

Material.

On separate pages of the booklet are drawn three lines. The first line, drawn vertically in the center of the page, is three inches long; the second line, similarly drawn on another page, is two and one-fourth inches long; and the third line, drawn horizontally, and having two curves, is three inches long. See Fig. IV.

Procedure.

"I want to see how well you can judge the length of lines without a ruler or even your pencils to help you measure. Look at this line carefully, and then draw a line under it here, horizontally, *i. e.*, from left to right like this, (hold pencil where line is to be drawn), which is 1) Just as long and no longer than this line.

"Now look at this next line. Draw a line under it, here, horizontally, which is 2) Just twice as long as this one.

"Let's play that this is a piece of string. If I could take hold of the ends of it, and pull the string out straight, how long a piece would it be? Draw a straight line here, to show how long it would be," (Point to space under the curved line).

Scoring.

The test is passed if two of the three lines are drawn within one-half inch of the required length, i. e.,

1) 3 in.
2) 4½ in.
3) 3 in.

Tests for Group IX.

IX, 1. Drinking Cup. (Bureau of Analysis, etc., N. Y.)[1]

Material.

A square of white paper, 7½ x 7½ in.

Procedure.

"I am going to show you how to make a drinking cup, and when I finish, I want you each to make one just like it. Watch, carefully.

1st: Fold the square through the center like this.

2nd: Take the right hand corner and fold it over to the opposite side so that these two edges are equal.

3rd: Next, take the other corner on the left side, and fold it over the same way so that the edges come straight across.

4th: Then, fold down this flap, and put it into this outer space to hold it.

5th: Now fold down the top flap for a cover, and you see how the cup looks.

You take your square of paper now, and make one."

Scoring.

The test is passed if the five folds of the drinking cup are correctly made.

IX, 2. Maze. (Porteus)[2]

Material.

The maze used by Porteus for nine year children, is printed in the booklet.

Procedure.

"Here you have a maze. Some of these paths are open, so you can get through, and some are closed at the ends. See how quickly you can find the shortest way out, without getting into any of the closed paths. Start where you see the letter *S*, and mark out the very shortest path you can find in order to get out over here on the other side. Do not stop to erase anything. If you make a mistake, just draw a line across it and go on as fast as you can."

1 New York State Board of Charities, Bureau of Analysis and Investigation: *Eleven Mental Tests Standardized*. 1915. Eugenics and Social Welfare Bulletin No. V. p. 42.

2. S. D. Porteus. *Mental Tests for Feeble Minded*. Journal of Psycho-Aesthenics Vol. 19, No. 4, June 1915, pp. 200–213.

Scoring.

The test is passed if the shortest path is marked out in one minute, with not more than two errors. An error consists of a line, however, short, drawn out from the shortest path.

IX, 3. Tapping Squares. (Whipple)

Material.

A sheet of paper is used on which a rectangle has been printed. This rectangle is divided into 150 half-inch squares, arranged in ten rows of fifteen squares each.

Procedure.

"Please take your pencils and hold them firmly about half way up. Place your arm comfortably on the desk so that it may bend freely at the elbow. Now when I say, 'Ready,' you are to tap as many squares as you can, in order, without hitting any square twice, or missing a square, or touching any of the lines. Work until I say 'Stop.'

Scoring.

The test is passed if 50 squares are correctly tapped in the 30 seconds allowed.

IX, 4. Alphabet Test. (Kuhlmann)

Procedure

"Of course you all know the letters of the alphabet, a-b-c-d-e, etc. Who can tell me what is the first letter before 'C'? (B). Now who can tell me what is the sixth letter before 'X'? (R). Let us put this one on the board, R-S-T-U-V-W-*X*. Notice that W would be the first letter before X, V the second, U the third, T the fourth, S the fifth and R the sixth.

"I will give you 30 seconds to answer each one of these questions.

1)	What is the 3rd letter before	K	(H)
2)	What is the 5th letter before	W	(R)
3)	What is the 2nd letter before	F	(D)
4)	What is the 1st letter before	Q	(P)
5)	What is the 4th letter before	E	(A)

Always write the letter to be used as a basis on the board, so that the children will not misunderstand.

Scoring.

The test is passed if three of the five letters are correct.

IX, 5. Memory Test. (Kuhlmann).

Procedure.

"I am going to write five words on the board, and let you look at them for a few seconds. Then I will erase them and write five others and let you look at them, and so on with the other series. When I am through, I will give you the first word of a series, and you are to write down from memory, the other four words that go with it. Then I will give you the first word of the next list, and you will write the four that go with that one, etc. You must not write down any until I tell you to."

Let them look at each series for 30 seconds, then erase and take the next. Give them 30 seconds to recall each list. The lists are:

2 G. M. Whipple. *Manual of Mental & Physical Tests*. 1910. pp. 100–115.

Silver	red	kitchen	blackboard	lawn
spoon	cow	stove	teacher	grass
table	fence	flour	desk	tree
bread	clover	water	slate	house
fork	dog	spoon	lesson	walk

Scoring.

The test is passed if the child gives three series without more than one error in each. An omission counts as an error.

Difficulties of Administering Group Tests. Many problems arise in giving group tests to the primary grades which, from their very nature, can never be entirely eliminated. Some of these difficulties are due to the distractibility of young children; to their inability to comprehend directions readily; to their undeveloped moral sense; to the attitude of the teacher; and to the apparent inhibition of response which the nervous child exhibits.

The inability of the young child to attend to set problems for any length of time is a serious difficulty. It necessitates devoting almost the entire first year of his school life to the formation of habits of work, of play, of thought and of living. Up to the time of his entrance into school, he is permitted to eat, sleep, and play largely at will, but with the beginning of his educational career he is expected to live according to a more or less definite plan, regardless of his own wishes or pleasure. When a stranger tries to test forty such children, before regular habits have been formed, the difficulties encountered may easily be imagined. Usually, the children are sufficiently shy and timid not to need disciplining, but sometimes that very timidity prevents them from being responsive at first. Therefore, their confidence must be won by the time the booklets are distributed and the examiner is ready to begin the testing. The fact that the children are not permitted to look through the pages except as they use them for the tests helps to keep up the interest. When the booklets are finally distributed, the directions for the first test given, the children attentively (?) awaiting the signal to begin work, and then a little voice in the corner confidentially informs one that "My sister has some gold beads just like yours," the examiner wonders if he can ever be sure of a child's attention.

The difficulty experienced by young children in comprehending directions has been discussed under the "Selection of Tests." Frequent illustrations, full and explicit directions with numerous repetitions, all help the child in understanding what he is to do.

However, ability to comprehend and follow directions accurately, is a test of intelligence, and too many or too great a variety of instructions may destroy the value of the test.

In order to stimulate the child to his best effort, one appeals to his competitive instinct, his sense of rivalry. This has its disadvantages, for the child in his eagerness to do better than his neighbor, or to get a word of praise from the examiner, resorts to methods "fair or foul." It is generally agreed that the young child is non-moral. The disapproval of the group has not been experienced sufficiently to make a lasting impression; and so the child slyly copies from his neighbor, or covertly writes on his desk the digits which are being read. To guard against such conditions as these, the writer has used various devices. Some have been mentioned under *Procedure*, such as having the child place his hands in his lap the instant he is given the signal to 'Stop'; covering his work with his hands as soon as he has finished the task assigned; or, in the digits test, place the pencils on the desks and leave them there until told to write. In one or two instances, where the attention had several times been called, without effect, to the fact that some-one in the room was not playing fair, more drastic measures have been taken, such as speaking to the child by name, or changing his seat. However, as most children are very sensitive to criticism, such measures are seldom necessary.

Frequently a teacher, not knowing the object of the group testing, feels that unless the pupils respond well to the tests, her reputation as a teacher will be at stake. In her anxiety to have the children make a good showing, she offers suggestions which help them. For example, when trying out various tests for the nine year group, the writer used Terman's test of finding all the words that rhyme with "Day, mill and spring." In one third grade, the teacher felt that the children were not responding as readily as they should, so with much irritation she spoke to the class. "Why, children, you know how we have been finding words to go in the 'ING' family, so I don't see why you can't find others like 'Day' to go in the 'AY' family." Of course the children could find words after that, but the test, as such, had lost its value. To avoid such incidents, the examiner should explain to the teacher at the outset that the purpose of the work is to see what each child can do by himself, without any suggestion from her.

So far, no difficulty has been experienced in securing the proper attitude on the part of the child, for the interest and novelty of the tests counterbalance any temperamental disturbances. However, it has been evident in the cases of several nervous children that the excitement produced by the competition, novelty and mental effort has inhibited responses, especially where the test has a time limit. If a pupil happens to be an only child and has not yet learned to adapt himself to an environment in which there are many children, he may fail to make an average score in the group tests, though when tested individually he may test even above normal.

Reliability. Increase in percentage passing for successive chronological ages. The discriminative capacity of group as well as of individual tests, is measured by the increase in the percentage of children passing the tests from one age to the next. The greater the increase in the percentage passing any test from one age to the next, the higher is the discriminative capacity of that test. Table II compares first, the percentage passing at age for each test in the five age groups, with the percentage of children, chronologically a year older, passing each test; and second, it compares the per cent. increase from one chronological age to the next as found in the group tests, with the per cent. increase found by Kuhlmann [1] and Terman [2] where they have used the same tests in their revisions as individual tests. For example, 66 per cent. of the six year old children passed the aesthetic comparison test (VI,1) where given in a group, and 75 per cent. of the seven-year olds passed it, thus giving an increase from one age to the next of 9 per cent. Where Kuhlmann used it as an individual test he found an increase of 4 per cent. passing from 6 to 7 years, and Terman found an increase of 6 per cent. at the same ages.

Table II.

Group V	Yr. 6 (68 cases)	Yr. 7 (79 cases)	% Increase	Indiv. % Increase Kuhlmann	Terman
1. Formboard	65	81	16	–	–
2. Counting 4 circles	63	81	18	1	–
3. Copying square	85	95	10	–	–
4. Discriminating Colors	78	81	3	–	9
5. Irreg. Tapping	70	85	15	–	–

1 These figures are taken from results in Dr. Kuhlmann's 1912 revision, in which the tests had not yet been correctly placed in the scale.

2 Terman's results as quoted are taken from the data published in 1917 on the Stanford Revision.

Group VI.	Yr. 6 (68 cases)	Yr. 7 (100 cases)			
1. Aesthetic Comparison	66	75	9	4	6
2. Mutilated Pictures	71	74	3	6	15
3. Irregular Taps	53	82	29	–	–
4. Two Commands	54	75	21	5	3
5. Perception of Sound	51	74	23	–	–
Group VII.	Yr. 7 (59 cases)	Yr. 8 (136 cases)			
1. No. of Toes	80	88	8	17	9
2. Memory Span	63	81	18	3	6
3. Copy Diamond	84	92	8	11	18
4. Tying Bowknot	68	73	5	–	16
5. Geom. Figures	61	76	15	–	–
Group VIII.	Yr. 8 (137 cases)	Yr. 9 (115 cases)			
1. Ball and Field	70	79	9	–	7
2. Dictation	78	87	9	–	8
3. Value of Stamps	70	79	9	20	17
4. Comprehension Test	72	78	6	–	–
5. Length of lines	69	81	12	–	–
Group IX.	Yr. 9 (59 cases)	Yr. 10 (90 cases)			
1. Drinking Cup	67	78	11	–	–
2. Maze	73	83	10	–	–
3. Tapping square	70	76	6	–	–
4. Alphabet Test	74	83	9	–	–
5. Memory Test	60	71	11	–	–

Unfortunately, there are no other similar group tests published with which the writer's percentages can be compared. However, there is about as much variation existing between the data from the two Revisions of the Binet Scale as between the individual and group results. In some instances, as in V, 4, Discriminating Colors, VI, 2, Mutilated Pictures, VII, 4, Tying Bow-knot, etc.—the tests show so little discriminative capacity as group tests, that the writer intends to find substitutes for them. Just why a test should show excellent discriminative capacity as an individual test and fail almost entirely when used in a group is at present a matter for conjecture to the writer. More data, however, will perhaps provide the explanation.

Considerable difference exists between the percentage of children passing at age in the group and in the individual tests. Dr. Kuhlmann has found that a much higher percentage of young children pass at age than of the older ones; thus, he found that 88 per cent. of the 4 year old children pass the 4 year tests, whereas only 54 per cent, of the twelve-year-olds pass the 12 year tests. However in the group tests, the writer finds as small a percentage of 6 and 7 year children passing 6 and 7 year tests, as of 9 year old children

passing the nine year tests. Table III compares the percentages passing at age as found by Dr. Kuhlmann [1] in the individual tests with those by the writer in the group tests. Thus, 78 per cent. of the six-year old children pass the six year individual tests, and only 60 per cent. of them pass the six year group tests.

Table III.

C.A.	*6* yrs.	*7* yrs.	*8* yrs.	*9* yrs.
Kuhlmann — Individual Tests	78	73	68	64
Lowell — Group Tests	60	69	72	68

The explanation of these differences in the six and seven year results may be found in the fact that a group test is largely a matter of comprehending and following directions presented orally, without any individual help for individual needs. If a child has not been used to doing things in a group, he will be easily distracted. Then, too, the younger the child the less capable he is of following instructions; but with each additional year of experience, and with an increasing mental age, the performance of tasks becomes easier. Therefore, the number of children passing a given test increases up to a certain point, with increasing chronological age. However, for ages eight and nine, the difference in percentage passing is possibly due to two causes, namely, the insufficient number of cases used in those age groups, and the questionable placing of some of the tests in the age groups. Further data are needed to corroborate this.

Increase in Number of Tests Passed for Successive Mental Ages.

If group tests are to be of value in selecting subnormal children, the number of tests passed in each age group by an individual must correlate with his age and intelligence. In the first three grades, used in the present survey, the chronological ages of the children ranged from six to eighteen years inclusive. The intelligence quotients of these children, determined after testing them with Dr. Kuhlmann's 1917 revision of the Binet tests, ranged from .40 — 1.30, though the largest number arrayed themselves between .70 — 1.00. In table IV, the average number of tests passed by the pupils of the various intelligence quotients, the average mental ages together with all the chronological ages found, are given for Grades I, II and III. Thus, in Grade I, chronologically six year old children, who had intelligence quotients from .70 — .79 inclusive, average

1 These figures were given me through the courtesy of Dr. Kuhlman, from his manuscript of *The Measurement of Mental Development, a further Extension and Revision of the Binet Simon Tests,* to be published soon.

4.7 years mentally and passed only 2.7 tests out of the age groups V, VI, VII; while those whose intelligence quotients were from .80 — .89. and whose mental age averaged 5.2 years, passed 4.3 tests, etc.

TABLE IV

Grade I (Groups V, VI, VII).

Chronological Age:[1] I. Q.		6 yrs.	7 yrs.	8 yrs.	9 yrs.	11 yrs.
.50– .59	Av. M. A.:	--	4.0	--	5.5	--
	No. Tests passed:	--	--	--	6.0	0 -
.60– .69	Av. M. A.:	--	4.4	--	--	7.3
	No. Tests passed:	--	1.9	--	--	7.9
.70– .79	Av. M. A.:	4.7	5.3	6.2	--	--
	No. Tests passed:	2.7	4.5	6.3	--	--
.80– .89	Av. M. A.:	5.3	5.9	6.6	7.5	--
	No. Tests passed:	4.3	5.7	7.5	9.5	--
.90– .99	Av. M. A.:	5.8	6.6	7.3	--	--
	No. Tests passed:	6.1	7.1	9.3	--	-0-
1.00–1.09	Av. M. A.:	6.2	7.1	--	--	--
	No. Tests passed:	8.3	9.8	--	--	--
1.10–1.19	Av. M. A.:	6.5	--	--	--	--
	No. Tests passed:	9.0	--	--	--	--

Grade II (Groups VI, VII, VIII)

Chronological Age: I. Q.		7 yrs.	8 yrs.	9 yrs.	10 yrs.	11 yrs.
.60– .69	Av. M. A.:	--	--	5.5	--	6.8
	No. Tests passed:	--	--	7.0	--	11.0
.70– .79	Av. M. A.:	--	6.3	6.4	7.3	8.2
	No. Tests passed:	--	7.4	8.8	11.5	12.0
.80– .89	Av. M. A.:	6.5	6.9	7.6	7.9	--
	No. Tests passed:	9.0	10.8	11.9	12.0	--
.90– .99	Av. M. A.:	7.2	7.5	8.1	8.9	--
	No. Tests passed:	10.0	11.0	12.3	13.0	--
1.00–1.09	Av. M. A.:	7.8	8.7	8.9	--	--
	No. Tests passed:	11.0	12.6	13.0	--	--
1.10–1.19	Av. M. A.:	--	9.3	--	--	--
	No. Tests passed:	--	14.0	--	--	--
1.20–1.29	Av. M. A.:	--	--	--	--	--
	No. Tests passed:	--	--	--	--	--

Since, as has been stated before, a group test in the primary grades consists largely of understanding and obeying instructions, and since these are also the fundamental factors involved in the first three years of school work, it is evident that the more years a child spends in those grades, the easier it will be for him to follow directions, or in other words, the more group tests he will pass. For example, Table IV shows that in Grade I children who are chronologically 7 years and whose average mental age is 7.1 — 7.5 pass 9.8 tests out of Groups V, VI, and VII; whereas children in

[1]Chronological age 6 includes ages from 5 yrs. 6 mos, to 6 yrs. 5 mos. inclusive; chronological age 7 includes ages between 6 yrs. 6 mos. and 7 yrs. 5 mos., etc.

Grade III, (Groups VII, VIII, IX.)

Chronological Age: I. Q.		8 yrs.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	15 yrs.
.40– .49	Av. M. A.:	– –	– –	– –	– –	– –	6.9
	No. Tests passed:	– –	– –	– –	– –	– –	6.3
.50– .59	Av. M. A.:	– –	– –	– –	– –	– –	8.1
	No. Tests passed:	– –	– –	– –	– –	– –	10.3
.60– .69	Av. M. A.:	– –	– –	– –	7.6	– –	9.4
	No. Tests passed:	– –	– –	– –	7.5	– –	13.3
.70– .79	Av. M. A.:	– –	7.0	7.7	8.3	8.8	–
	No. Tests passed:	– –	9.0	10.4	8.0	10.3	-
.80– .89	Av. M. A.:	6.8	7.8	8.5	8.8	9.6	–
	No. Tests passed:	8.5	10.6	12.2	9.0	11.3	–
.90– .99	Av. M. A.:	7.7	8.4	9.0	– –	– –	–
	No. Tests passed:	9 5	12.0	13.5	– –	– –	–
1.00–1.09	Av. M. A.:	8.7	– –	9.6	– –	– –	–
	No. Tests passed:	10.0	– –	14.0	– –	– –	–
1.10–1.19	Av. M. A.:	9.2	9.5	– –	– –	– –	–
	No. Tests passed:	12.0	15.3	– –	– –	– –	–

Grade II, with the same chronological age and average mental age, but who have been in school two years, pass 10 tests out of Groups VI, VII and VIII; and in Grade III children of the same chronological and average mental ages pass 12 tests out of Groups VII, VIII and IX. This increase in the number of tests passed, although the tests are of greater difficulty for each succeeding grade, seems to be accounted for by the additional years of school experience.

The use of the group tests in the survey has been to sort out the subnormal children, or those whose mentality is questionable, for further individual testing. If more complete standardization of the tests confirms the present results, it will be possible to make such a classification very readily. For example, upon giving the Group Test, V, VI and VII to a first grade, one could compare the number of tests passed by any child of a given chronological age with the number of tests passed by children of the same grade and chronological age, but with varying intelligence quotients, and when the numbers correspond the intelligence quotients can be approximated. Thus, if an eight year old child in the first grade passes only 6.3 tests, one may be fairly certain that his intelligence quotient will be somewhere between .70 — .79. Since this places him in the "borderline" group, he will undoubtedly need individual testing before his mental status can be correctly gauged.

COMMUNICATIONS AND DISCUSSIONS

A SCALE IN UNITED STATES HISTORY

The article in this JOURNAL for May, 1917, by Bell and McCollum on "A Study of the Attainments of Pupils in United States History" presents very obvious possibilities of making a much needed scale in that subject. To this end the writer has made certain calculations from the tabulated data in the article and has now re-issued the whole series of tests in convenient form with directions for giving the tests, scale values for each of the questions, directions for scoring, and standards for various school groups. The original data were obtained by Bell and McCollum with the use of the following test material in a large number of groups of students ranging from the fifth grade of the elementary school to the senior class in the University.

Test I. Give the reason for the historic importance of each of the ten representative dates: (1) 1861. (2) 1789. (3) 1620. (4) 1565. (5) 1898. (6) 1619. (7) 1783. (8) 1492. (9) 1776. (10) 1846.

Test II. Indicate for what each of these prominent men was celebrated: (1) John Burgoyne. (2) Alexander Hamilton. (3) Jefferson Davis. (4) Walter Raleigh (5) John C. Calhoun. (6) Cyrus H. McCormick. (7) George Dewey. (8) Sam Houston. (9) Roger Williams. (10) James Oglethorpe.

Test III. Mention the name of the man prominently connected with each of these historic events: (1) Captured Quebec during French and Indian War. (2) Discovered the North Pole. (3) Wrote the Declaration of Independence. (4) Invented the telephone. (5) Brought about the Missouri Compromise. (6) Captured the city of Mexico during the Mexican War. (7) Founded the Colony of Maryland. (8) Made a great speech against the English Stamp Act. (9) Was President of the United States during the Civil War. (10) Vetoed the re-chartering of the United States Bank.

Test IV. Define in a short sentence each of the following historic terms: (1) Second Continental Congress. (2) Lewis and Clark Expedition. (3) Articles of Confederation. (4) Sherman Anti-Trust Law. (5) Monroe Doctrine. (6) Fugitive Slave Law. (7) Dred Scott Decision. (8) Alien and Sedition Laws. (9) Nullification Ordinance of South Carolina. (10) Emancipation Proclamation.

Test V. Make a list of all the political parties that have arisen in the United States since the Revolution, and state one principle advocated by each.

Test VI. Indicate the great divisions or epochs of United States History.

Test VII. In the accompanying outline map of the United States draw the land boundaries of the United States at the close of the Revolution and indicate the different acquisitions of territory since that date. (Outline map 4 x 5 inches supplied).

The object of this report is to give some details of the derivation of the scale values and of the subsequent standard scales. The plan here adopted follows closely that employed by the writer in his

article "A Scale in Ancient History," published in the same number of the JOURNAL though there are some distinct variations and necessarily some limitations due to the fact that the original data was not obtainable and the published tables alone must be used. It will be observed from Table I of the Bell and McCollum article that forty-one groups of subjects were employed. The size of each group varied from eight to eighty-six and the total comprised 1476 individuals. Grades for these groups are presented in percentage of correct answers for each of the forty-three questions, except that there are no elementary school records for tests V, VI, and VII. Individual grades not being available, the nearest approximation to the distribution of those grades is the distribution of their group averages. Five calculations were made from the grades on representative questions and the Standard Deviations were found to be 17, 18, 18, 21, and 22. An average of these gives 19, a close approximation to the standard deviation for the various groups on any of the questions. In the following calculations, this is taken as the "Sigma" value. Its use is justified by rather bold assumptions but in spite of the apparent skew in the curve, its value is probably as well established as the nature of the data warrants.

The most representative set of grades available is that for 1425 Normal School, High School and Elementary School percentages shown in Table III, p. 270, of Bell and McCollum's study. Supplementing this by the weighted average grades of these people on tests V, VI, and VII taken from Table I, p. 264, achievement grades of 18%, 21%, and 38% respectively are obtained for this same group. These are shown in the "grade" column in Table I arranged as they were in the original test.

The median of these forty-three grades is found to be 28%. From this the "deviations" of the second column are obtained. Now, dividing each of these differentials by sigma (19), indices of the distance above or below the median of difficulty are found for each question. These indices are shown in the third column for each question, and bear reverse signs for the reason that the data is in terms of achievement while the subsequent scale values must be in terms of difficulty. To show this, the respective position on the distribution curve must be reversed. When the zero point on the base line is obtained, it must be a zero of difficulty of questions distributed along that base line, rather than a zero of ability to answer questions.

The method of approximating zero difficulty was by means of difficulty interpolation beyond the range in terms of range units of difficulty. This, of course, assumes that the group is fully representative. For instance, if the averages of the three extremes in each case are taken to determine range limits, they are found to be ten per cent. and seventy-nine per cent. respectively. This shows a range span of 69 per cent. The median deviations of these

TABLE I

Showing percental achievement grade, median deviations, index quotients of standard deviation distance on the base line, and derived scale values referred to an interpolated zero.

	Grade	Dev.	Index	Sc. Val.		Grade	Dev.	Index	Sc. Val.
I– 1	44	+16	–84	294	III–1	35	+7	–37	341
2	17	–11	+58	436	2	75	+47	–247	131
3	27	–1	+15	383	3	50	+22	–116	262
4	17	–11	+58	436	4	15	–13	+68	446
5	35	+7	–37	341	5	17	–11	+58	436
6	17	–11	+58	436	6	15	–13	+68	446
7	21	–7	+37	415	7	20	–8	+42	420
8	89	+61	–321	57	8	26	–2	+11	389
9	53	+25	–131	247	9	60	+32	–168	210
10	22	–6	+32	410	10	12	–16	+84	462
II– 1	29	+1	–5	373	IV–1	21	–7	+37	415
2	28	0	0	378	2	32	+4	–21	357
3	45	+17	–89	289	3	25	–3	+16	394
4	54	+26	–14	364	4	14	–14	+74	452
5	39	+11	–58	320	5	35	+7	–37	341
6	16	–12	+63	441	6	32	+4	–21	357
7	43	+15	–79	299	7	21	–7	+37	415
8	73	+45	–237	141	8	12	–16	+84	462
9	33	+5	–26	352	9	7	–21	+81	459
10	35	+7	–37	341	10	26	–2	+11	389
V–	18	–10	+53	431 (0)					
VI–	21	–7	+37	415 (0)					
VII–	38	+10	–53	325 (0)					

range limits are –18 and +51 respectively, which when divided by sigma, give indices of 94 and –268. Thus the span 69 per cent. is equal to the sum of these values, or 362 units. The upper limit of achievement, 79 per cent. is 21 per cent. from zero difficulty. Calculating from the above, zero is found to be 378 units from the median. Other spans of extreme and mean ranges were used in the calculation, but as the results differed only in fractions from this result, this was taken as a proper determination. Assuming, then, that zero difficulty is 378 units below the median of difficulty, and combining 378 algebraically with the values in column three, Table I, in this study, the final scale values of column four are obtained. In actual practice these are expressed in abridged form by moving the decimal point back to the left one place, except in the case of tests V, VI, and VII, where the grade is made up from a single measure, rather than from ten separate measures. This discrimination gives proper weight to these latter tests in the final score of the pupil.

In the scale as now made available in convenient form for use in schools these values are inserted beside each question for convenience of grading and the table of standards (Table II) is included.

TABLE II.

Showing total values for each test, achievement standards for four school groups, and sex differences for each of the seven tests.

	No.	Test I	TestII	TestIII	Test IV		Test VI	Test VII	Tot.
Value	1425	347	329	360	435	431	415	325	2669
University	75	153	179	126	174	142	265	146	1184
Normal School	207	149	176	171	152	134	203	107	1091
High School	668	90	121	106	97	109	141	181	745
El. School	550	37	63	46	25	X	X	X	171
Boys	579	128	141	133	109	134	145	117	907
Girls	846	111	115	101	91	100	146	85	740

These standards were obtained by turning the scale values back upon the percental achievements as distributed through Tables I, III and IV of Bell and McCollum's article.

The advantage of the scale over the original method of grading lies in the fact that it gives proper weight to the achievement of answering any one question. In the original study each question was graded as of equal difficulty with every other one. This was obviously not justified as a means of ranking pupils but was proper as a means of securing data on which to base a scale of difficulty of questions. In the standards here furnished, proper weighting is given to the answer for each question and the totals represent the probable achievement of each group. The method of grading is to credit the pupil with the score equal to the scale value of the question when it is accurately and adequately answered. In the case of partial answers the scale value should be appropriately apportioned according to the merit of the answer as measured by an absolute standard, rather than by one's idea of what a child of that particular grade might be expected to answer. Complete instructions are given with the test material itself.

In closing, the writer can announce the appearance in the near future of still another series of tests in Unites States History along quite divergent lines from this one. Instead of measuring the range of information in only a few aspects of the subject, the new series will more nearly cover the range of what history teaching should accomplish, and will likewise more nearly eliminate the subjective factor in grading.

The University of Texas. L. W. SACKETT.

CORRELATION BETWEEN TRANSLATION OF LATIN PROSE AND ETHICAL DISCRIMINATION[1]

The data for this study were obtained from twenty six juniors and seniors in the Ethical Culture High School, of New York City. They were given the following tests in the spring of 1918.

Latin Test.

Translate at sight the following text and write out your translation in good clean English, representing the thought of the author as carefully as possible. (The passage given was taken from Cicero and was so selected as to give a fair test of first-hand translation, but with familiar vocabulary).

Ethical Test.

Write out the answers to the following ethical problems according to your best judgment.

1. A throws a snowball through the school window. B sees him do it. When the teacher asks who broke the window, A remains silent. What should B do about it? Describe as many different points of view as possible and support the one you think right.
2. A is riding on a street car and the conductor fails to collect his fare. What should A do about it? Defend the position you take by giving as many reasons as you can think of.
3. A steals $5.00 and B steals $5,000. Which is the more guilty of doing wrong? Defend the position you take and give all the reasons you can.
4. A owes B $5.00. One day they meet at the cashier's window of a bank. As A turns away with his money he drops a $5.00 bill. Has B the right to keep the $5.00 which he sees A drop in payment of what A owes him? Defend the position you take with reasons.
5. Suppose that the day before their meeting at the bank B had asked A for the money on the street and A had refused to pay him and had abused him roughly for asking for the money. Under these circumstances would B have a right to keep the money he sees A drop. Defend the position you take with reasons.

---

[1]Pedagogical Studies in the Ethical Culture School No. 1. This study was made by the author while he was pedagogical physchologist in the Ethical Culture School, New York City.

No. of Pupil	Rank in Latin	Rank in Ethics	No. of Pupil	Rank in Latin	Rank in Ethics
1	1.0	22.5	14	13.0	13.0
2	2.5	17.5	15	13.0	22.5
3	2.5	21.5	16	17.5	1.0
4	5.0	5.0	17	17.5	2.5
5	5.0	8.5	18	17.5	5.0
6	5.0	13.0	19	17.5	24.5
7	8.5	17.5	20	21.0	2.5
8	8.5	21.5	21	21.0	8.5
9	8.5	24.5	22	21.0	17.5
10	8.5	26.0	23	23.5	8.5
11	13.0	8.5	24	23.5	17.5
12	13.0	13.0	25	15.0	5.0
13	13.0	13.0	26	26.0	13.0

Both sets of papers were gone over very carefully and sorted into ten piles each, from the best to the poorest, on the basis of keenness of discrimination manifested in the translation of the Latin and in the analysis of the ethical situations presented.

The following is a summary of the rank comparisons obtained: The above comparisons yields a correlation of –.62, when calculated by the formula—Rho equals one minus six times the sum of the rank differences squared, divided by the number of cases times the quantity, the number of cases squared minus one. Since the form of distribution is approximately that of the normal probability curve, the correlation may be inferred from the obtained value of Rho by the use of table No. 36, p. 168, *Mental and Social Measurements*, by Thorndike.

If the formula R equals one minus six times the sum of the plus difference in ranks, divided by the number of cases squared minus one is used and the correlation value inferred from table No. 37 in Thorndike, the result is –.93.

ROBERT A. CUMMINS

Simpson College, Iowa.

TIME ASSIGNMENT AND RATE OF IMPROVEMENT IN HANDWRITING

There follow a description of the method and a statement of the results of an experiment conducted in grades six A, seven B, and seven A, in three of the white elementary schools of Guthrie, Okla., from January to May, 1915, covering a period of four and a half months, for the purpose of determining the effect upon improvement in handwriting of different amounts of time spent in practice.

Method.

The pupils of each school spent in class practice periods fifty (50), seventy-five (75), and one hundred (100) minutes per week, respectively. No attention was given to handwriting other than incidental attention in connection with all written work. There had always been a definite attempt on the part of all teachers concerned with instruction of classes under investigation to secure as good handwriting as possible in all written work, so that in this experiment no extra attention was given to that feature by the teachers of the fifty-minute and seventy-five minute groups.

The first week in January in the middle of the regular penmanship period three copies were made of a short verse which all the children of a room knew, and the best one was picked out; *i. e.* the one the teacher, without reference to any measuring scale, thought the best of the three. Then at the end of the training series, the middle of May, the same procedure was followed, using the same verse, the pupils not knowing at any time that any experiment was being carried out.

These samples of handwriting were all graded or measured the middle of May by the principals of these elementary schools, according to both the Freeman scale and the Ayres scale. One person graded the sixth A specimens of all three schools, both the January and the May specimens; one person, likewise, graded all the seventh B specimens, and one person, all the seventh A specimens. In this way we eliminated any discrepancy that might be due to different judges measuring or grading the work of pupils in some grades and different schools, and different judges measuring beginning and final tests of the same pupil. Beginning papers of the whole group were graded and then final papers, so that the judges, in grading final papers, would have no idea of any particular pupil's score on the beginning paper.

These results were tabulated, and from them the following tables were made, showing the amount of improvement, positive and negative, in each grade in each school, and the number of pupils making each amount of improvement. No attempt has been made to bring out all the possible interpretations of the data here presented, but only some of the general and tentative conclusions.

RESULTS

In considering results I will confine myself to measurement by the Freeman scale because it, by analyzing handwriting into its elements, makes possible more accurate measurement and diagnosis.

TABLE I

Amounts of improvement and number of cases of each amount in each grade and each school

Improvement	6	5	4	3	2	1	0	1	2	3	4	5	6	7	8	9	10	Total cases
Seven A.																		
School A.	0	0	0	0	0	0	2	0	3	4	0	1	0	0	1	0	0	11
School B.	0	0	0	0	1	2	5	4	5	5	0	1	3	1	2	0	1	30
School C.	0	0	0	0	1	2	1	0	2	6	2	4	2	1	1	0	0	22
Seven B.																		
School A.	0	0	0	1	0	4	3	5	4	11	3	1	1	2	0	0	0	35
School B.	0	0	0	0	1	0	4	4	1	4	1	1	2	0	1	0	0	19
School C.	1	0	0	0	1	0	1	2	1	2	1	2	0	0	1	0	0	13
Six A.																		
School A.	0	0	0	1	0	0	1	3	4	4	0	1	1	0	0	1	0	16
School B.	0	1	1	1	2	2	2	1	3	1	1	2	2	0	0	0	0	21
School C.	0	0	1	0	0	4	4	3	0	3	1	0	1	0	0	0	0	17
Totals(7A, 7B, 6A).																		
School A.	0	0	0	2	0	4	6	8	11	19	3	3	2	2	1	1	0	62
School B.	0	1	1	1	4	4	11	11	7	12	3	4	6	1	3	0	1	70
School C.	1	0	1	1	1	6	6	5	3	12	4	6	4	1	1	0	0	52
Minutes per week—School A, 50; School B, 75; School C, 100.																		

Comparing improvement of the different practice-period groups and showing per cent of whole group making different amounts of improvement we have

TABLE II.

Improvement	Zero & Neg.	1–3	4–6	7–10	Total	
School A.	19.35%	61.29%	12.90%	6.45%	99.99%	. .
School B.	31.43	42.85	18.57	7.14	99.99	. .
School C.	30.77	38.46	26.92	3.84	99.99	. .

The weighted average amounts of improvement of the three groups of pupils were as follows:—

School A, 2.37 (Median 2.52)
School B, 2.08 (Median 1.85)
School C, 2.13 (Median 2.70)

These figures show that the fifty-minute group made as a whole the most improvement. The hundred-minute group had, however, more cases of marked improvement. The negative improvement may be accounted for by the following:—1st. Some pupils, entering the schools that year for the first time had bad writing habits (especially bad movement) to be broken up and there was loss while this was being done. Shifting of school population was greatest in School B, and least in School C. 2nd. The stage of physical development of some pupils was undergoing rapid changes, new coordinations, etc., being established, so that there was a noticeable lack of mascular control. This is borne out by the fact that, while the six A group (ages 12-13) constituted but 29.34% of the three groups, 48.14% of the negative improvement was in this grade. 3rd. In some cases there seemed to be a lack of interest in penmanship improvement, and carelessness and inattention resulting from this, led to poorer work. School B had to contend with this condition the most, and School C, the least. If one eliminated negative improvement on the grounds that it (in the first two cases above) is a temporary condition, the weighted average improvement would then be as follows:— School A, 2.53

School B, 2.42
School C, 2.53

In view of the necessity for economy of time in education it is worth while that more extensive investigations, covering a longer period of time, and including many more pupils and a wider grade distribution of pupils, be made as to the effect of time assignment upon the rate of improvement. This experiment, covering four and a half months time, and including 184 pupils in grades six and seven indicates that (where ordinary care in handwriting is required in all written work) the amount of improvement is not proportional to the amount of time spent in practice; that the fifty-minute-per-week assignment to handwriting practice in these grades gives as good results as the seventy-five-minute or the hundred-minute assignment.

Fowler D. Brooks.
State Normal School, Mankato, Minn.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive reseaiches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3 00 pei year of nine monthly issues, payable in advance. Single issues are 50 cents each Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X SEPTEMBER NUMBER 7

EDITORIAL

COMPOSITION SCALES FOR THOUGHT CONTENT

Undoubtedly the objective in most composition writing is good form of expression. The teacher's attention is fixed, not on what the pupil thinks, but on the way in which he expresses his thought. Some teachers, however, are shrewd enough to realize that the chief difficulty lies not in the mere form of expression, but in the fact that the pupil has no thoughts to express. These teachers insist that the real task of composition work is thought building. Instead of refining and polishing and deleting and rearranging, the composition teacher should take everything that is presented and call for more, – – always more. True, the pupil should be encouraged to recognize when he is thinking vigorously and richly, should be asked to discriminate between sturdy, luxuriant, copious thinking and lean, thin, halting paucity of ideas. The only way to do this, however, is to stimulate him to think, and that is chiefly an affair of subject matter, – – of content rather than of form. To teach composition;

then, you should let the pupil imagine that you care little or nothing about compositions as such, but that you are tremendously interested in some topic in which he is interested, and that you want to know everything about it that he can possibly find out and tell you.

To what extent do our composition scales help us to measure improvement in this sort of thing? Most of them are composed of detached and isolated paragraphs on topics in which a child would have only an artificial interest. But is it not possible to reduce to more exact terms any improvement in thought content? The teacher is well aware that in any class there are tremendous individual differences in thought content on any topic that may be proposed. The difficulty in trying to develop a scale to measure these differences is that one would have to have a different scale for each unit of subject matter that might be proposed for pupils' reflection. But perhaps the task is not so hopeless as it seems. If we should choose ten topics in which children are interested and on which some children think well, and should give each of them to a sufficiently large number of children, we might get such evenly distributed variations of thought content as would lend themselves to the divisions of a scale. On this basis we could construct ten scales based on representative types of thought content.

In dealing with any given pupil, John Smith, we could determine by a series of exercises his ability to think along each of these lines, and thus could construct a fairly accurate 'profile' of his thinking power. If he were high in some of these fields and low in others, the question would then arise whether it were better to try to develop thinking power along the lines in which he is now weakest, and thus attain what might be called 'all-round' thinking ability, or whether it would be more advantageous to cultivate intensively those fields in which he is already somewhat proficient. Further would come the question as to what type of exercise is best fitted to develop thinking power in a given field, what is John's individual learning capacity in the field in question, to what extent does his improvement in this field carry over into other fields, etc. There would undoubtedly be technical difficulties encountered in constructing such scales, but their value for diagnostic purposes and for studies in the art of teaching composition would be so great as to amply repay any effort that might be expended on them. With adequate control of the attitude of the writer toward the task (by asking him to place himself under definite conditions) important contributions might be made to the study of the development of the higher thought processes.

J. C. B.

NOTES AND NEWS

Dr. David Spence Hill, of the department of education, University of Illinois, has been elected president of the University of New Mexico.

Dr. W. W. Charters, dean of the school of education of the University of Illinois, has become professor of education and director of the research bureau for retail training at the Carnegie Institute of Technology. Dr. J. B. Miner, who has been acting director of this bureau, will continue with the bureau and will also give advanced courses in vocational psychology.

Dr. Grace E. Bird, psychologist at the Rhode Island Normal School, has been appointed professor of educational psychology at the Rhode Island State College. These two institutions are now affiliated for the purpose of giving degrees in education.

Dr. Clinton P. McCord, health director of the Albany, N. Y. schools, has been appointed head of the newly created bureau of health in the Pennsylvania State Department of Education.

The London Education Committee has recommended that Mr. Cyril Burt's position as psychologist should be made permanent.

Dr. George O. Ferguson, Jr., professor of psychology and education at Colgate University, has been appointed professor of education at the University of Virginia.

Professor Truman L. Kelley, who was in charge of the administration of the plan of the committee on education and special training for the rating and classification of soldiers in the S. A. T. C., has resumed his work in Teachers College, Columbia University.

At Teachers College, Columbia University, Franklin W. Johnson, principal of the University of Chicago high school has been appointed associate professor of education, Dr. Leta S. Hollingsworth has been promoted from instructor in educational psychology to assistant professor of education, and Dr. William A. McCall has been advanced from instructor in experimental education to assistant professor of education.

Dr. J. V. Breitwieser, professor of psychology and education at Colorado College, has been appointed associate professor of education in the University of California.

Dr. W. S. Foster, Assistant professor of psychology at Cornell University, has been appointed professor of psychology at the University of Minnesota.

Dr. Harry D. Kitson, instructor in psychology at the University of Chicago, has accepted the position in psychology at Indiana University made vacant by the election of Professor E. C. Lindley to the presidency of the University of Idaho.

VOLUME X OCTOBER 1919 NUMBER 8

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

THE SPELLING ABILITY OF COLLEGE STUDENTS

ELSIE MURRAY

Sweet Briar College, Sweet Briar, Virginia

I. PROBLEM

The apparent decline in spelling ability in the present generation of college students has been subject to various interpretations by parents and instructors. The latter are inclined to cast the blame upon lowered standards of instruction and achievement in the schools; the former, to minimize spelling as on the whole an inferior, and, in the present era, a somewhat superfluous accomplishment. The student himself is prone to shift the responsibility for his shortcomings upon some extraneous factor, associated with the educational system to which he has been subjected; *e. g.*, forced speed in note-taking, and the attendant deterioration of handwriting and spelling together; the study of foreign languages, and the resultant confusion of English words with those of similar origin in other tongues; current interest in phonetic spelling, and the like. Excuse is also sometimes offered on the basis of purely intrinsic peculiarities; a strain of bad spelling hereditary in the family (a Mendelian unit characteristic, as it were); a natural distaste for this 'branch of learning'; etc.......

To the scientific inquirer, various other explanatory factors suggest themselves: the possible role, under current methods of instruction, of a defective equipment along the lines of visual imagery; the probably influence (in many cases of so-called 'hereditary bad spelling') of 'social heredity,' *i. e.*, the perpetuation of parental shortcomings through family correspondence (of especial importance in this era when every decade sees a smaller proportion of college freshmen recruited from families in which the use of correct English is a tradition). There remains a considerable residue of apparently

inexplicable cases; and to these especially the interest of the writer has been attracted by the circumstance that the four most aggravated cases of bad spelling personally encountered in a number of years have resolved themselves upon examination into instances of defective hearing, vision, articulation, and of an excessively nervous temperament, respectively. Additional observations have suggested the possibility of a definite correlation between spelling capacity in the individual and other abilities, *viz.*, accuracy in certain branches; clearness or mistiness of ideas; likelihood of success or failure in various college subjects. Hence, while relegating the improvement of methods of instruction to pedagogical science, the psychologist may find to his hand a problem of no less interest in the investigation of the intrinsic causes of poor spelling and their significance to the individual.

Apart from these theoretical interests, the following study was intended to serve a practical and pedagogical purpose. For some time, the writer had entertained a suspicion that the blame for the semi-illiterate spelling of many of our graduates rests less upon the defects of early instruction than upon the irrational attitude of college instructor and student; an attitude, on the one side, of exasperation and scorn; on the other, of happy-go-lucky indifference, pessimism, fatalism— 'once a poor speller always a poor speller.' One immediate aim was, therefore, to break up this pessimism of the student, to create in its place an attitude of self-criticism, an interest in the orthographic integrity of the word as the outward symbol of the idea. With this in view, the college bulletin board was pressed into service, graphs representative of scores, tabulation of results for different classes, improvement, etc., posted. An attempt was also made to modify the instructor's attitude by dwelling on the fact that the acquisition of an adult spelling vocabulary cannot reasonably be relegated to the grades, but should (and with a little attention does) run parallel to the development of the corresponding concepts throughout the college course. Secondarily the study served also the purpose of illustrating to students in education modern methods of attacking and presenting pedagogical and related problems.

II. Material, Methods, Subjects.

List A. The first test consisted of 20 words, selected at random from a list of words commonly misspelled by students in philoscphy and psychology, compiled by the writer over a number of years.

In order to avoid misunderstandings (such as the mistaking of *rhythm* for *ribbon*), the enunciation of each word in the test was followed immediately by reading an illustrative phrase. (This phrase was not written down by the student). The list comprised the words: *rhythm, judgment, definite, parallel, embarrass, repetition, nucleus, separate, reverence, hereditary, paralysed, omitted, tendency, analyse, efficient, primitive, mathematics, sensible, explanatory, disappointed.* Approximately 3 minutes was consumed in the test.

All members of the four college classes, 204 in all, were given this test, in small sections, in the fall of 1917. Since a number of the students tested declared that, while unable now to spell a number of these words, they had been able to spell them all when in the grades, their spelling having deteriorated while in college, it was thought of interest to try the test upon the seventh and eighth grades in the public schools, 284 pupils of which were available. The test was also given to 69 members of library training classes.

Six months later the junior and senior classes were retested with the same list.

List B. This test was made up on somewhat different lines. Twenty words were selected from the published lists of Kallom and others, representing as far as possible the typical difficulties of English spelling. The list included *precipitated, occurrence, mortgage, twelfth, principal, proceeded, mischievous, marriageable, preceded, achievement, superintendent, chimneys, height, receipt, desiccated, relieved, principle, all right, privilege, misspelled.* These words were incorporated in a series of sentences, which were cut up into 24 phrases, and read at a uniform rate, with a pause after each phrase, the whole consuming about 6 minutes.

The student was not informed that her spelling was being tested, but was merely told that a dictation exercise would be given, at a fairly rapid rate; that strict attention must be paid, as no portion would be repeated; that no abbreviations could be used, and all i's must be dotted and t's crossed; that no alterations should be made in a word once set down; and lastly that the sense of the selection must be noted in order to ensure the writing of the correct word. These instructions were formulated in order to secure a measure of the normal spelling ability of the student, by approximating as far as possible the conditions of spontaneous writing, in which the cue is afforded by verbal imagery (visual or otherwise), and in which the attention is not unduly centered upon the mechanics of spelling.

With this end in view no special effort was made to enunciate each syllable with absolute distinctness, and the speed kept up was such that as a rule the subject must write the crucial words in the pauses between the phrases, from auditory after images. Further, the prohibition against the altering of words once set down was designed not so much to obviate the possibility of correction from a neighbor's copy, as to avoid a source of confusion revealed in the earlier test; in which the well-known law of attention (according to which any complex figure, such as a word, on being submitted to inspection for a given time, tends to break up into new parts, and take on a strange aspect) came into operation, with the result that many words at first correctly written were afterwards erroneously altered by the student in a fit of self-consciousness.

This test was given the Seniors and Juniors only, in the fall of 1917, and repeated in the spring of 1918, six months later.

III. Results. What Do The Test Scores Mean?

List A. Distribution of Scores. Their Significance. (a) *Superiority of Upper Class Scores.* There is not only a progressive increase (though slight) in the number of words spelled correctly from the freshman to the junior year (see Table I), but also a decided superiority of the college over the seventh and eighth grade scores. This

TABLE I. LIST A. QUARTILE DISTRIBUTION OF SCORES

	Grades		College			
	Seventh	Eighth	Freshmen	Sophomore	Junior	Senior
Best	16	19	20	20	20(20)	20(20)
75%ile **	12	14	18	19	19(20)	19(20)
Median	*9*	*11.5*	*16*	*17*	*18*(19)	*18*(19)
25%ile	7	9	14	15	17(18)	17(18)
Worst	1	4	9	7	14(15)	14(10)
Number tested	167	122	77	54	37	36

**The 75 percentile in this table is the score attained or exceeded by 25% of the group; fallen short of by 75%.

increase may, of course, be interpreted not as indicative of improvement in spelling ability with each successive year in college, but merely as a by-product of selective forces operating from college entrance to the senior year; a view apparently supported by the fact that in the junior year, when elimination practically ceases, the spelling scores also become stationary. Since, however, elimination

is usually on the basis of scholarship, and spelling ability and scholarship by no means run parallel (see below), it seems justifiable to overlook the selective factor, and interpret the rise in scores as at least controverting the student claim that spelling necessarily deteriorates in the course of a college career, if not, indeed, indicating a tendency toward positive improvement.

b) *Improvement in Scores on Repetition of the Test.* To test further the susceptibility to improvement in spelling during the college course, the test was repeated after the lapse of six months, with the two upper classes. No warning of this repetition was given, though in the meantime the scores for the four college classes had been posted, along with a copy in script * * * of the twenty words, in the order of the frequency of their misspelling. In the case of the seniors, the subject of spelling in general, and of improved methods of instruction had been discussed in a class in education.

The resulting scores (see Table I in parentheses) show a general improvement, though there are numerous irregularities, reflected in the relatively low correlation between the first and second set of records, +.50 for the senior class, +.60 for the junior. It is interesting to note that a few students maintained that the first exercise was to a certain extent disorganizing, made them over-selfconscious in their spelling, and created an attitude of uncertainty. As a matter of fact, in the group of 73 juniors and seniors, three originally perfect scores fell off one point on second trial, four others dropped slightly, while the score of one Senior fell from 14 to 10. It seems reasonable, however, to interpret this merely as a general instability in spelling arcs or images (indicated even in the original test by numerous alterations in the written word), rather than as a positive deterioration.

Further, personal knowledge of the individuals inclines the writer to the opinion that the individual differences in amount of improvement indicate not so much difference in degrees of improvability, as differences in incentive to improvement. No special appeal for improvement was made in connection with the test, but students with known ambitions along cultural or vocational lines undoubtedly made the greatest effort. The results may, therefore be taken to indicate what may be accomplished with a proportionately small expenditure of time and attention.

***Script was adopted on the theory that the uncertainty of students in the spelling of their adult vocabulary may be due in part to lack of clear visual images *in script* of words which they have many times encountered in print.

c) *Discrepancies between Scores* and *Self-Estimates.* Each subject, before submitting to the spelling test, was requested to rate herself as a *good, poor,* or *medium,* speller. The resulting estimates are interesting, first of all, in view of the variations in the percentage of '*Good*' and '*Poor*' judgments given by members of each class (see Table IIa). The proportion of '*Good*' decreases steadily from seventh grade to the sophomore class, where it undergoes a sudden inflation, to fall again through the senior and junior years. Whether or not these fluctuations represent characteristic tendencies toward over and underestimation in general will afford interesting matter for further experimentation.

TABLE IIa. DISTRIBUTION OF SELF ESTIMATES (*Percentage*)

	GRADES		COLLEGE			
	Seventh	Eighth	Freshmen	Sophomore	Junior	Senior
GOOD	43%	35%	23%	44%	30%	25%
MEDIUM	48	55	64	41	43	58
POOR	9	10	13	15	27	17

As to the approximate reliability of these subjective estimates, it may be noted that for the four college classes the number of misspellings in the individual scores averages twice as great for those estimating themselves as 'Poor' as for those estimating themselves as 'Good.' (See Table IIb) At the same time, in the group of 204

TABLE IIb. RELATION OF SELF ESTIMATES AND SCORES

Self Estimate	Freshmen	Sophomore	Junior	Senior		Total Average
Good	17.4	16.9	18.1	18.6	Average Score	
	2.6	3.1	1.9	1.4	Average Error	2.2
Poor	12.9	12.3	16.8	17	Average Score	
	7.1	7.7	3.2	3	Average Error	5.2

college students, of the 62 estimating themselves as good, only 8 (or approximately 13%) made a perfect score; although 43 of the 62 (about 70%) scored above the median for the College (17). The most striking cases of discrepancy are tabulated in Table IIc below.

The more extreme cases of over-estimation cited above (*i. e.*, of students rating themselves as 'Good' and scoring low in the test) would seem to indicate, as we had anticipated, the need of establishing a more critical attitude toward self-accomplishment. On the

TABLE IIc. OVER AND UNDER ESTIMATIONS

		OVERESTIMATIONS		UNDERESTIMATIONS	
Individuals whose self-estimate = 'Good'				= 'Poor'	
Class	Median Score	Number	Scores	Number	Scores
Seventh	9	2	4 and 5	1	16
Eighth	11.5	2	8 and 9	1	15
Freshman	16	2	12 and 15	1	18
Sophomore	17	2	13 and 14	1	17
Junior	18	4	16 and 17	1	19
Senior	18	1	17	2	19

other hand, the cases of under-estimation (*i. e.*, of students rating themselves as 'Poor' and scoring fairly high in our test), while relatively few, and possibly explicable as instances of exaggerated self-depreciation, tend to cast some doubt upon the reliability of our test as a measure of spelling ability. This doubt is further substantiated by the skewing of the curve of distribution for the college scores toward the upper end of the scale; a fairly certain indication of the use as measuring rod of a test too easy for the group. (See Table III).

TABLE III DISTRIBUTION OF SCORES (LIST A) FOR THREE GROUPS

Scores	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
GRADES	4	1	3	5	12	21	18	30	31	42	23	28	27	15	16	5	4	3		
LIBRARY						2				1	1	1	6	9	8	7	9	13	10	2
COLLEGE							1		2	1	5	4	8	14	17	22	45	24	46	15

d) *Value of List A as a Measure of the Spelling Ability of the College Student.* There are four possible criteria of the value of such a list; comparison of the records, first, with actual school grades in spelling; second, with degree of success in a standardized exercise in checking misspelled words; third, with self estimates; fourth, with the scores of a second spelling test. The first method was feasible only with the seventh and eighth grades, and, since it seemed hardly likely that conclusions arrived at on the basis of data from younger students would be valid for older ones, was abandoned. The second yielded a low correlation (+.40 for seniors, +.39 for juniors). Since, however, experiment has shown that ability to reproduce and to recognize do not necessarily run parallel, this evidence was also finally rejected as negative. The third criterion, comparison with self-estimates, upon closer scrutiny, guided by a knowledge of individual cases, appeared to be too much complicated by subjective factors, over-confidence and over-modesty, to stand by itself. Hence it seemed advisable to devise a second test, which, if not representing a fairer sample of the student's writing vocabulary, should at least present more of the fundamental difficulties of the English language.

TABLE IV a. FREQUENCIES. LIST A. AND B. SCORES FOR 37 SENIORS AND 37 JUNIORS.

Scores=		6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
List A	Junior									3		4	8	6	11	5
	Senior									3	1	3	10	7	8	5
List B	Junior	1				1	1	4	6	2	5	5	6	4	2	
	Senior	1		1		2	1	1	3	7	6	5	3	6	1	

TABLE IV b. QUARTILE DISTRIBUTION OF SCORES

	Juniors		Seniors	
Test	A	B	A	B
Best	20	19(19)	20	19(19)*
75%ile	19	17(17)	19	17(18)
Median	18	15(16)	18	15(17)
25%ile	17	13(14)	17	14(16)
Worst	14	6 (9)	14	6 (9)

B. *List B. Scores and Their Significance*

a) *Distribution of Scores*

The median score for both juniors and seniors is lower than in Test A (15 as against 18—see Table IV b), and there is a wider range between the limits (6-19 as against 14-20 for Test A). That is, the test is more difficult, brings out a wider range of individual differences, and therefore gives greater promise of gauging the spelling ability of the group. The junior and senior records again closely parallel each other, corroborating the findings of Test A, and indicating cessation of progress with the junior class.

b) *Improvement on Repetition.* This test also was repeated six months later, less attention, however, having been paid in the meantime to the list, which had merely been written once on the blackboard for each class. There is, again, a general improvement in the scores for the second trial, showing that many took the lesson of the first one to heart, though no special effort was made to impress it. The degree of improvement is greater for the senior than the junior class, the median score rising from 15 to 16 for the latter, to 17 for the former; probably for the reasons already indicated, and perhaps also because of the more serious attitude of the senior as the close of her college career approaches, and the problem of vocational qualifications begins to assume importance.

*The figures in parentheses represent the results of the second trial.

The correlation between the individual scores for trials 1 and 2 is slightly higher than for List A (+.62 for Seniors, +.65 for Juniors). There are, however, as in the case of the earlier test many irregularities in the degree of improvement, due probably in part to inequality of effort and interest, in part to chance factors. It is, for instance, discouraging to note that in the case of the two Seniors registering the greatest gain in the second trial in List A, the score of the one remains stationary on retrial in List B, while that of the other drops. In all, in the group of 74, there are 14 cases of lowered scores, as against 8 in the retrial of List A, (see Table IV c). The greatest drop, however, is of 3 points only (1 case); while 17 records register gains of 3 to 5 points, and the total number of misspelled words is reduced from 360 to 283. Of those registering 3 points improvement, 6 were individuals estimating their spelling ability as 'Poor', thus demonstrating the value of even a minimum of scientific attention to the spelling of the adult. A more systematic direction of effort would probably have brought up the scores of the 6 'Poor' cases whose scores either remained stationary or dropped.

c) *Correspondence between Self Estimates and Scores.* Here as in Test A the average number of errors scored by those estimating

TABLE V a.

	List A			List B			
Self Estimate	Junior	Senior	Av.	Junior	Senior	Av.	
Good	1.9	1.4	*1.6*	4.4	3.2	*3.8*	Average Error
	18.1	18.6		15.6	16.8		Average Score
Poor	3.2	3	*3.1*	7.6	7.3	*7.4*	Average Error
	16.8	17		12.4	12.7		Average Score

their ability as 'Poor' is nearly twice that scored by those estimating themselves as 'Good' (see Table V a.); while the average score of the 'Poor' falls as before in the lowest quartile of the group records. Also, as before, there is a slightly closer correspondence between estimates and test scores in the case of the seniors than of the juniors.

While there is no striking difference in the figures for the two tests as shown in Table V a, it will be seen by reference to Table V b that Test B is effective in reducing the cases of most glaring discrepancy between estimates and scores from 8 to 4, eliminating them entirely in the case of the seniors.

TABLE Vb.

Test A	
No. Juniors estimating themselves as 'Good' who score below median	=4
No. Seniors estimating themselves as 'Good' who score below median	=2
No. Juniors estimating themselves as 'Poor' who score above median	=1
No. Seniors estimating themselves as 'Poor' who score above median	=1
	8=Total
Test B	
No. Juniors estimating themselves as 'Good' who score below median	=3
No. Seniors estimating themselves as 'Good' who score below median	=0
No. Juniors estimating themselves as 'Poor' who score above median	=1
No. Seniors estimating themselves as 'Poor' who score above median	=0
	4=Total

d) *Reliability of Test B as a Measure of Spelling Ability.* The correlation of scores with misspelled word checking records is lower for List B than for List A; +.34 for juniors, +.38 for seniors. Combination of the results from the two lists, however, raises these coefficients to +.40 and +.54, whatever the significance of this fact may be (see above). The evidence furnished by the comparison of self estimates and scores is positive, indicating the greater reliability of the second list (see paragraph c). The low correlation of the scores for the two tests (+.50 for Seniors, +.42 for Juniors) is susceptible of various explanations. It is possible that it is due in part to a difference in method, the second test calling for stricter concentration, and putting a higher premium upon speed and poise. It undoubtedly also indicates the lower reliability of one of the tests, or of either taken singly. Therefore, in what follows the combined scores of the two tests are used as probably furnishing a more reliable measure of spelling ability than either taken singly.

IV. Correlation of Spelling Ability with Other Abilities

a) *Spelling Ability and Accuracy in General.* The correlation between rank in the spelling tests and in a single addition test (Thorndike's single-digit short-column test) is low (+.29)*; while that between spelling and grading in freshman mathematics is even lower, averaging +.12½ for juniors and seniors. It is, further, worthy of note throughout the results that individuals specializing in mathematics in the junior or senior year almost invariably rank low in spelling.

*Calculated for juniors only.

At the same time, the correlation between the spelling scores and a composite accuracy score ** is fairly high, averaging +.65. Since the corresponding correlation with a composite reasoning tests score (based on ten tests in hard opposites, logical memory, hard directions, mixed relations, definitions, paragraph completion, word building, proverb matching, information, classification of concepts) is lower, +.59, and that with academic grades for three years (seniors only) only +.42, the evidence seems to indicate that spelling ability possesses a closer relation to the factors making for accuracy than it does to those making for general intelligence or academic ability.

TABLE VI.

CORRELATION COEFFICIENTS (*Method of Rank Differences*)

	Misspelled Word	Addition Test	Accuracy Tests	Reasoning Tests	Definition Test	
Av. Spelling Scores	+.40		+.62	+.57	+.62	Senior
" " "	+.54	+.29	+.72	.	+.54	Junior
	+.47		+.67		+.59	Average

	College Grades	Mathematics	Latin	English	
Av. Spelling Scores	+.42	+.13	+.44	+.59	Senior
		+.12	+.23	+.52	Junior
		+.12½	+.33½	+.5½	Average

b) *Spelling Ability and Standing in College Courses.*

English. In correcting a definitions test (based on the Terman and Kirkpatrick vocabulary lists), observation of the confusion to which the student mind is subject in the case of words of similar sound or appearance (such as chagrin and shagreen, fin and fen, ochre and ogre) gave rise to the suggestion that the vagueness of verbal images evident in the one test might correlate high with the vagueness of orthographic images in the other; the ability to manip-

**Based on ten tests in the case of Seniors—checking misspelled words, easy directions, digit and group digit cancellation, handwriting pairing, spelling, handwriting, learning, immediate memory, addition; on the first six only in the case of juniors.

ulate the outward form (spelling) with that to manipulate the inner core or meaning. As a matter of fact, the average correlation coefficient between the spelling and definitions tests for the two upper classes is +.58; between spelling and grades in freshman English, +.59. The correlation with the reasoning tests is also +.59; but since the latter were of a type largely dependent on verbal facility no less than thought processes it may perhaps be fairly assumed that spelling is a correlate not only of a general tendency toward accuracy, but also though perhaps in a lesser degree, of the subject's interest in English as a medium of culture and self-expression; that clear-cut verbal concepts to a certain extent presuppose firmly fixed spelling neurograms. There are, of course, numerous exceptions. Moreover, individuals may undoubtedly be letter perfect in spelling as a mechanical exercise, and possess only the vaguest concepts of the meanings of words they spell.

Latin The close dependence of spelling ability upon familiarity with Latin roots has been strongly argued by more than one classical instructor. It is, therefore, worthy of remark that the correlation between the spelling scores and grades in freshman Latin,* while higher than that for freshman mathematics (+.12½), is nevertheless low enough to be negligible (+.33½).

V. Causes of Misspelling

a. *General.*

Words most frequently misspelled. The relative difficulty of the words in the two lists used is indicated by the Tables VIIa and b, which give the per cent. of errors scored for each word. An attempt has been made also to analyse the sources of trouble in the forty words used, and to estimate their relative importance (see Table VIII). While the classification adopted is admittedly unsatisfactory, and the captions selected not always mutually exclusive, the scheme presented serves the purpose of emphasizing some of the typical difficulties on which attention needs to be centred in any constructive effort to improve the spelling of the individual.* The first four heads of the classification (covering more than half of the words in the lists, and including those on the whole most subject to error) represent, of course, difficulties intrinsic to the unphonetic character

*Calculated for seniors and juniors only.

Note* The construction of the two lists was, of course, too casual to admit of any sweeping generalisations on the basis of Table VIII as to the most important sources of spelling difficulty either in the language at large, or in the vocabulary of the college student.

TABLE VIIa.

PERCENTAGE FREQUENCY OF MISSPELLINGS. LIST A. 204 COLLEGE STUDENTS

	List A	Per Cent Misspellings	Most Common Error
1.	embarrass	56.3*	embarass
2.	paralysed	42.3	paralised
3.	rhythm	36.7	rythm
4.	judgment	35.8	judgement
5.	repetition	24	repitition
6.	analyse	23	analise
7.	nucleus	22	neucleus
8.	hereditary	10.7	heriditary
9.	primitive	10.3	primative
10.	definite	9.3	definate
11.	reverence	7.9	reverance
12.	explanatory	6.9	explanitory
	tendency	6.9	tendancy
13.	mathematics	6.4	mathamatics
	parallel	6.4	paralell
14.	disappointed	5.9	dissapointed
15.	separate	5.4	seperate
	efficient	5.4	effecient
16.	omitted	3.4	ommitted
17.	sensible	1.5	sensable

TABLE VIIb.

PERCENTAGE FREQUENCY OF MISSPELLINGS. LIST B. 74 COLLEGE JUNIORS & SENIORS

	List B.	Per Cent Misspellings	Most Common Error
1.	desiccated	98.6	dessicated
2.	all right	54	alright
3.	preceded	48.6	preceeded
4.	occurrence	35	occurence
5.	marriageable	27	marriagiable
6.	privilege	25.7	privelege or priviledge
7.	misspelled	21.6	mispelled
7.	superintendent	21.6	superintendant
8.	mischievous	20.2	mischeivous
9.	mortgage	17.5	morgage
10.	chimneys	16.2	chimnies
10.	twelfth	16.2	twelth
11.	precipitated	14.8	percipitated
12.	achievement	13.5	achievment or acheivment
12.	proceeded	13.5	proceded
12.	height	13.5	heighth
13.	principal	10.8	principle
14.	principle	9.4	principal
15.	relieved	8.1	releived
16.	receipt	6.7	receit

*I. e., approximately 56 out of every 100 college students may be expected to misspell the word *embarrass*.

of our mother tongue. The remaining three heads represent difficulties assignable rather to 'psychological' or individual and remediable causes. For while it might be argued that the seventh difficulty has its real origin in the excessive variability of our language in its ways and means of representing practically identical sounds (see heading 1), it seems equally valid to attribute it to lack of finesse in the training of the ear and the mechanism of articulation (thus practically identifying it with heading 6). The median liability to error is only slightly higher for the first four classes of words division 1)—14.9 per cent.—than for the last three (division II)—10.7 per cent.—hence it seems probable that considerable improvement in spelling may be effected without having recourse to spelling reform.

b. *Individual.*

Relaton of Defective Vision, Hearing, Articulation, etc., to Spelling Ability. In order to investigate more minutely the possible causes of good and bad spelling, each of the 74 juniors and seniors was requested to fill out a questionaire, the fifteen or more heads of which referred in part to amount and character of early training, in part to preferences in school studies, in part to corrected or uncorrected defects in vision and hearing. The more significant data contained in the answers of the twenty worst and the twenty best spellers have been tabulated and summarized (along with certain facts gathered from other sources) in Tables IX A, B and C. Table IX A, read from left to right, shows the *average spelling score* of the subject, her *own estimate* of her spelling ability, whether or not the majority of her *family* are *good spellers,* whether she *liked* or *disliked spelling* as a school subject, whether she was drilled on a *speller* or *selected lists* of words in the grades, whether the study was continued into the *High School,* whether her *vision, hearing,* and *articulation,* are good or not, her *handwriting legible,* her *temperament* high-strung, even, or lazy, her *academic work* above or below par, her *preference* for literary mathematical, historical, scientific, classical, or musical studies.

Examination of these Tables would seem to corroborate the view that individual peculiarities (more especially defective hearing, vision, or articulation) form the most formidable handicap to the acquisition of habits of correct spelling. At the same time it must be admitted that there is no one decisive factor, but in practically every case a number of coordinate and interlocking causes.

TABLE VIII TYPICAL ERRORS AND THEIR FREQUENCY. LISTS A AND B.

	CLASSIFICATION		FREQUENCY	NUMBER OF WORDS
I. 1.	Substitution of similar sounding vowels and diphthongs:	paralise	42.3%	5
		analise	23	
	i for y	mischeivous	20.2	
	ei and i	releived	8.1	
		acheivement	6.7	
			Median = 20.2	
2.	Doubled letters:	dessicated	98.6	8
		embarass	56.3	
		occurence	35	
		mispelled	21.6	
		proceded	6.7	
		paralell	6.4	
		disapointed	5.9	
		ommitted	3.4	
			Median = 14.1	
3.	Analogy:	alright (always)	54	9
		preceeded (proceeded)	48.6	
		judgement (acknowledgement)	35.8	
		chimnies (berries)	16.2	
		heighth (breadth)	13.5	
		priviledge (knowledge)	12.8	
		principle	10.8	
		principal	9.4	
		preceeded (for proceeded)	6.7	
			Median = 13.5	
4.	Silent letters:	rythm	36.7	4
		morgage	17.5	
		achievment	6.7	
		receit	6.7	
			Median = 12.1	Total = 26 (24)
II. 5.	Influence of other syllables in the same word:	repitition	24	4
		neucleus	22	
		heriditary	10.7	
		effecient	5.4	
			Median = 16.3	
6.	Faulty pronunciation:	twelth	16.2	5
		percipitated	14.8	
		heighth	13.5	
		heriditary	10.7	
		effecient	5.4	
			Median = 13.5	
7.	Unaccented syllables (substitution of similar sounding vowels):			
	e and i	repitition	24	11
		privelege	12	
	a and i	primative	10.9	
		definate	9.3	
		explanitory	6.9	
		sensable	1.5	
	a and e	superintendant	21.6	
		reverance	7.9	
		tendancy	6.9	
		mathamatics	6.4	
		seperate	5.4	
			Median = 7.9	Total = 20 (17)

TABLE IX A.

INDIVIDUAL CASES OF THE TWENTY BEST SPELLERS IN THE JUNIOR AND SENIOR CLASSES

Subj.	Class	Spell Score	Self Est.	Fam.	Lik.	Meth.	H. S.	Hear.	Vis.	Art.	Hwt.	Temp.	Aca. Gr.	Subject Most	Subject Least	Eng.
1.	S	19.5	M	G	Yes	Sp.	No	G	G	G	G	Even	M	Lit.	Math.	G
2.	S	19.5	G	G	Yes	Sp.	No	G	G	G	M	Even	G	Lat.	Sci.	G
3.	J	19.	M	G	Yes	Sel.	Yes	G	G	G–	G	Nerv.	M	Math.	Sci.	P
4.	J	19.	M	G	Yes	Sel.	No	G	M	G	G	Laz.	M	Lit.	Math.	G
5.	J	19.	M	G	Yes	Sp.	No	M	P	G	G	Even	G	Math.	Art	G
6.	S	18.5	G	G	Yes	Sp.	Yes	G	G	G	G	Even	G	Mus. & Lit.	Lat.	G
7.	S	18.5	G	G	Ind.	Sp.	No	M	M	M	G	Even	G	Lit.	Math.	G
8.	S	18.5	G	G	Yes	Sp.	No	G	P	G	P	Nerv.	M	Lit.	Math.	G
9.	J	18.5	G	M	Yes	Sp.	No	G	G	G	G	Even	G	Hist.	Lat.	M
10.	J	18.5	M	G	Yes	Sp.	No	M	M	G	M	Nerv.	P	Mus. & Lit.	Lat.	P
11.	S	18	G	G	Yes	Sp.	No	G	M	G	M	Nerv.	P	Lit.	Math.	M
12.	S	18	G	G	Yes	Mix.	No	G	M	G	M	Nerv.	P	Lit.	Math.	M
13.	S	18	M	G	Ind.	Sp.	Yes	G	G	G	G	Even	G	Lit.	Math.	M
14.	S	18	M	P	Yes	Sp.	Yes	M	M	G	M	Even	M	Lit.	Sci.	M
15.	J	18	M	M	Yes	Sp.	Yes	G	G	G	G	Nerv.	G	Mus. & Lit.	Math.	G
16.	J	18	M	G	Yes	Sp.	No	M	G	G	P	Even	G	Lit.	Sci.	G
17.	J	18	G	G	Yes	Mix.	Yes	G	M	G	M	Even	G	Lat.	Mus.	M
18.	J	18	G	G	Yes	Sp.	Yes	G	M	G	G	Even	M	Lit.	Math.	P
19.	J	18	G	G	Yes	Mix.	No	G	M	G	G	Even	M	Lit.	Math.	M
20.	J	18	G	G	Yes	Mix.	Yes	G	M	G	P	Even	M	Lit.	Lat.	M

In the above table, G = Good, M = Medium, P = Poor, Ind. = Indifferent.

TABLE IX B.

INDIVIDUAL CASES OF THE TWENTY WORST SPELLERS IN THE JUNIOR AND SENIOR CLASSES

Subj.	Class	Spell. Score	Self Est.	Fam.	Lik.	Meth.	H.S.	Hear	Vis.	Art.	Hwt.	Temp.	Aca. Gr.	Subject Most	Subject Least	Eng.
1.	J	15 3	G	G	Yes	Sp.	No	G	M	G	P	Even	M	Hist.	Lat.	P
2.	S	15	M	M	No	Mix.	Yes	G	G?	G	P	Nerv.	G	Mus.	Math.	G
3.	S	15	M	G	Ind.	Sp.	No	G	G	G	G	Lazy	M	Sci.	Lat.	G
4.	S	15	M	M	No	Sp.	No	G	M	G	G	Even	M	Sci.	Hist.	G
5.	J	15	M	M	Yes	Mix.	No	M	G	G	P	Nerv.	M	Hist.	Math.	P
6.	J	15	P	M	No	Sp.	No	G	P	G	P	Nerv.	M	Lit.	Lat.	P
7.	J	15	M	M	Yes	Sp.	No	P	P	G	P	Nerv.	M	Lit.	Lat.	P
8.	S	14.5	P	P	No	Sp.	No	G	G	G	M	Even	M	Math.	Lat.	P
9.	S	14.5	M	G	Yes	Sp.	No	G	M	G	P	Lazy	P	Lit.	Math.	P
10.	S	14.5	M	G	Yes	Mix.	No	M	P	G	P	Nerv.	P	Hist.	Sci.	P
11.	J	14.5	G	P	Yes	Sp.	No	G?	P	G	G	Laz.	P	Hist.	Math.	P
12.	J	14.5	P	M	No	Sp.	No	G	G	G	M	Even	M	Hist.	Math.	M
13.	J	14.5	G	M	Yes	Sp.	Yes	G	P	G	M	Even	M	Mus.	Math.	P
14.	S	14	M	M	Yes	Mix.	No	M	M	G	G	Nerv.	M	Lit.	Math.	M
15.	J	13	P	P	No	Mix.	No	G	M	M	G	Lazy	M	Sci.	Hist.	P
16.	J	12.5	P	M	No	Sp.	No	G	G	G	P	Lazy	M	Sci.	Art	P
17.	J	12.5	P	M	No	Sp.	No	G?	M	G	P	Lazy	M	Sci.	Lat.	P
18.	S	12	M	G	No	Sp.	No	G	G	P	M	Even	P	Hist.	Lat.	P
19.	S	11	P	G	No	Sel.	No	M	P	M	P	Nerv.	M	Hist.	Sci.	P
20.	S	10.5	M	G	Yes	Sp.	No	G	P	M	M	Lazy	P	Math.	Hist.	P

TABLE IX c.
SUMMARY OF CASES OF TWENTY BEST AND TWENTY WORST SPELLERS

LIKING	Yes	Indif.	No.	
	18	2		Best Spellers
	9	1	10	Worst Spellers
METHOD	Speller	Mixed	Selected	
	14	4	2	Best Spellers
	14	5	1	Worst Spellers
HIGH SCHOOL	Yes		No	
	8		12	Best Spellers
	2		18	Worst Spellers
TEMPERAMENT	Even	Lazy	Nervous	
	13	1	6	Best Spellers
	6	7	7	Worst Spellers
HEARING	Good	Medium	Poor	
	15	5		Best Spellers
	15	4	1	Worst Spellers
VISION	9	10	1	Best Spellers
	7	6	7	Worst Spellers
ARTICULATION	19	1		Best Spellers
	16	3	1	Worst Spellers
HANDWRITING	12	5	3	Best Spellers
	5	5	10	Worst Spellers
ACADEMIC GRADES	10	8	2	Best Spellers
	1	14	5	Worst Spellers
FRESHMAN ENGLISH	9	8	3	Best Spellers
	3	3	14	Worst Spellers
FAMILY	17	2	1	Best Spellers
	7	10	3	Worst Spellers
SUBJECTS LIKED	Most		Least	
	Literature (11)		Mathematics (10)	Best Spellers
	History (7)		Mathematics (7)	Worst Spellers

1. *Early Training.* There is no clear evidence that it matters one way or another whether the individual has been drilled on the basis of the old-fashioned speller, or brought up on more modern methods, with a selected word list, derived from the day's work. Prolongation of study into the High School, on the other hand, seems to exert a definitely favorable influence; while in the case of certain individuals (notably numbers 9 and 14 of the worst spellers) irregularities in early schooling, such as shortening of the term in the grades, private tutoring, country schools, seem to have been the chief determining factors of deficiency. The fact that 15 per cent. only of the good spellers (as against 65 per cent. of the poor spellers) come from homes in which incorrect spelling is said to be the rule, may

also be considered here, as an indication that bad spelling in one's early home environment is as contagious as bad grammar. The figures would probably be even more striking, if a more reliable estimate of the spelling in the 'poor spellers' families were available.

2. *General Learning Ability.* Ability to spell (as measured by our tests) and general learning ability (as gauged by academic grades) are by no means commensurate, as has been already indicated by reference to their low correlation coefficient (+.42). Such correlation as there is, however, is most evident in the upper and lower reaches of the scale (for spelling ability); 50 per cent. of the best spellers ranking also among the best students, whereas only 5 per cent. of the worst spellers fall in this class. The correlation is somewhat less obvious near the foot of the scale, only 25 per cent. of the poorest spellers falling in the rank of the poorest students, while 10 per cent of the best spellers belong here also. On the whole, however, it will perhaps be found most satisfactory to accept general learning ability as the basic factor in correct spelling, explaining the exceptions and anomalies by reference to peculiarities of early training, of the individual's sensory or motor equipment, or interests (See headings 1, 3 and 4).

3. *Sensory and Motor Peculiarities.* Seventy-five per cent.* of the poor spellers have some defect or other of hearing, vision, or articulation, (one half of them a decided defect, and nearly as many a defect in more than one of these respects), as against sixty-five per cent. of the good spellers, not one sixth of whom have more than a minor deficiency. Most of the visual cases among the former are cases of near-sightedness or astigmatism, either uncorrected in the grades, or corrected with glasses worn with great irregularity in school or college. The bearing of this latter defect upon exactness of orthography may be readily surmised in the light of the fact that for the astigmatic eye black characters upon white seem to dance upon the page and blur, steady fixation or exploration of minute details producing headache which may amount to virtual torture. Vagueness of auditory images, in cases of imperfect hearing, imperfectness of articulatory imagery, where there is an impediment of speech, may be assumed to offer an equal handicap, even when (in the latter case) the articulatory defect does not directly affect the portion of the word most liable to error. In the opinion of the writer, all three conditions conduce to inexactness.

*The writer suspects that this figure would be even higher if careful eye tests were made on all subjects, notably cases 2 and 3 of the worst spellers, who have virtually nothing against them on our chart as it stands except temperamental peculiarities.

The significance of the fact that 50 per cent. of the poorest spellers (as against 15 per cent of the best) fall in the class of the least legible writers is not so clear. Illegible handwriting may be a contributory cause (of verbal inexactness), or merely a coordinate effect of some more or less obscure factor making for sensory-motor inexactness in general. The same may be said with regard to peculiarities of temperament or disposition.

4. *Interests.* The possibility that handicaps in the way of sensory, motor, or temperamental defects may be effectively counterbalanced by intensive direction of the subject's interest to literary fields is indicated by the cases of 7, 8, 10, and 14, of the best spellers (Table IX A), and by the fact that, on retrial, the scores of those among the poor spellers of the first trial who rated literary studies high almost uniformly took an upward leap. The fact that 55 per cent. of the best spellers declare a preference for literary studies, while 50 per cent. place mathematics at the foot of the list, is significant in this connection,* as are also the correlation coefficients between spelling, and grades in English and mathematics, quoted earlier (Table VI a). These figures would, to be sure, gain in significance if corroborated by the inverse location of these subjects on the lists of preferred studies of 'poor spellers' instead of which we find history at the head, though in no clear majority.

As might perhaps be anticipated, 45 per cent. of the best spellers (as against 15 per cent. of the poorest) took high rank in freshman English; while 15 per cent. of the former, 70 per cent. of the latter, scored with the lowest.

The early liking for spelling itself as a school subject, found most frequently among good spellers (in 90 per cent of cases), seems to be correlated with, or subsidiary to, interest in literary studies in general. In any case, its importance is ambigious, since it is as likely to represent a result as a cause of spelling ability; *i. e.*, it is dubious whether the individual does well in a given subject because he possesses natural ability in that direction and enjoys working in it; or whether, his interests having been directed that way by chance (*e. g.*, by social pressure, or family culture) special effort is exerted along the lines of this subject, and liking follows success or vice versa.

---

*It is interesting to note at this point the statement frequently made by grade teachers to the effect that children who do well in spelling and reading usually do poorly in mathematics and vice versa.

VI. Summary

1. There is no evidence of deterioration in the spelling of college students from year to year.

2. Although early habits of error are difficult to overcome, the spelling of adults may be improved with a minimum of effort; indeed, it seems highly probable that the college course represents the most fitting moment for the acquisition of spelling arcs in connection with the majority of the words of the adult vocabulary, *per se*.

3. Correction of sensory defects in early childhood would probably prove as efficacious as improved methods of teaching in eliminating faulty spelling.

4. Direction of interests toward or away from literary studies strongly affects spelling. Interest in mathematics appears to exert an inhibitory influence. Family influences probably operate indirectly here, as reinforcing factors or the reverse, rather than directly through the medium of heredity, *i. e.*, in the opinion of the writer, the 'born speller' (and his opposite) is a myth.

5. Correlation between spelling ability and general ability, while positive, is not so striking, as that between spelling ability and certain factors making for general accuracy and exactness.

'BRIGHT' AND 'SLOW' PUPILS IN ELEMENTARY AND HIGH SCHOOL*

ROBERT A. CUMMINS, PH.D.†
Louisiana State Normal School, Natchitoches, La.

The investigation here reported set out to answer the following questions:

(1) What per cent. of "Bright" pupils in the high school were "Slow" in the "Early" grades?

(2) What per cent. of "Slow" pupils in the high school were "Bright" in the "Early" grades?

(3) If such cases are found, what explanation can be offered for such change in the standing of these pupils?

(4) Other minor questions, such as may arise in connection with the study, with reference to progress and elimination.

Before proceeding with the investigation it was first necessary to define the terms "Bright" and "Slow," and to set the limitations of the terms "Early" and "Late," as applied to entrance to school. For the purpose of our study, the dividing line between "Early" and "Late" entrance was placed at the end of the fifth grade. Thus all the pupils who entered in the fifth grade or earlier were classed as "Early" entrants and all who entered in the sixth grade or later were classed as "Late" entrants.

It was decided to class the upper quartile of each group investigated as "Bright" and to class the lower quartile as "Slow." This involved the adoption of some system of ranking the pupils in respect to their scholarship. The only source of information available consisted of a set of high school record cards, on which were recorded the marks given by the teachers and a set of elementary record cards containing the teachers' estimate of the pupils. The high school marks were reported six times each year for each subject taken and in some cases twice for each subject, as for example, English, for which two grades were sometimes recorded, one being for composition and the other for literature. Two grades were sometimes found in the case of Latin and a few other subjects. The

* A study of the records of 116 graduates and 184 non-graduates for the years 1913-14 to 1916-17, inclusive, who attended the Ethical Culture School, of New York City.

† Sometime pedagogical psychologist to the Ethical Culture School, of New Yorl City.

system of marking used was as follows: A equals very good—90 to 100%; B equals good—75 to 90%; C equals passing—60 to 75%; C— equals not passing—condition; D equals failure. Effort is shown by the use of a plus mark and lack of effort by the use of a minus mark. (In the case of C— lack of effort is shown by an additional minus mark.) The elementary record cards were regularly handed in several times each year and more often as occasion might require. For a pupil who entered the kindergarten and continued on through the high school to graduation, there were as many as 150 to 200 of these cards on file. While these are called elementary record cards, it appears that similar cards are turned in by the high school teachers, so that the complete report for each pupil from the time of entering the school to the time of leaving is found in one file, with the exception of the *marks* received in high school. There were also found incomplete grade cards for the years below the high school. There appeared to be an inverse correlation between the scholarship rating of any given pupil and the number of report cards turned in, *i. e.*, the higher the standing, the fewer the number of cards and the lower the standing, the more cards were turned in by the various teachers all the way through the grades and the high school.

As a basis for ranking the pupils it was decided to use a method of weighting the marks received in high school and check this rating by comparing it with the judgment of the teachers as found on the report cards. The following formula was devised for the evaluation of the marks received in high school:

$$\text{S. Q.} = \frac{3A+2B+C-D}{A+B+C+D} \text{ in which}$$

A, B, C, and D represent the number of each mark received, respectively, and S. Q. represents the 'Scholastic Quotient' sought. This quotient has a range of from minus 1 to plus 3. In the accompanying table a few of the more representative values are shown in terms of the relative proportion of each mark necessary to yield the given quotient. The highest possible quotient, of course, is when all the marks received are A's and the lowest possible quotient is when all the marks are D's. The number of marks received does not alter these limits, since the same result is obtained with one mark or with a thousand and one. There are numerous possible combinations of

Limits of range of Q.'s are—1.00 & 3.00

3	00	Al	A			
2	50	Equal No.	A	B		
2	00	Equal No.	A	B	C	
2	00	All		B		
2	00	Equal No.	A		C	
1	50	Equal No.		B	C	
1	25	Equal No.	A	B	C	D
1	00	Equal No.	A			D
1	00	All			C	
	67	Equal No.		B	C	D
	50	Equal No.		B		D
0	00	Equal No.			C	D
— 1	00	All				D

the four marks that would produce various degrees of the quotient, ranging theoretically from the highest, or plus 3.00 to the lowest, or minus 1.00.

It should be pointed out in this connection that the use of this formula assumes an equal weight for all the subjects for which marks are given. For this reason we should expect some cases in which, for example, a pupil would be placed in the class of "Bright" high school pupils because of a high S. Q. and at the same time be rated very much lower by the combined judgment of the teachers. This discrepancy would be explained by the fact that the most of the high marks received might conceivably be in subjects such as art, shop, or physical training, which are not reckoned as contributing as much toward scholarship in this particular school, as do the other subjects. Such a case is that of girl No. 6 of the class of 1914-15, who shows the very high S. Q. of 160. This places her in the upper half of the

middle 50 per cent of a group of 64 graduates, who entered "Late." As a matter of fact, however, 11 of the 15 A's received were given in the subject of physical training.

This raises the question of what subjects furnish the best criterion for estimating scholarship and hence the best basis on which to predict success or failure in life. The answer to this question very naturally depends upon the kind of success in life one has in mind. If success in life means to become a successful teacher of Latin or Greek, or any other language for that matter, manifestly the only persons who will ever be successful are those who master the languages which are to be taught. So, also, if success in life means to attain a high degree of prominence in some vocation, which involves the use of higher mathematics, manifestly those who are to be counted as successful must master higher mathematics. Or, again, if success in life is interpreted to mean the passing of college entrance examinations and the completion of a classical course in some institution of higher learning, it goes without saying that success in life can be predicted by success or failure in the same subjects in the high school. But, if it is allowed that a millionaire broker, or manager of a great manufacturing concern, or a manager of a big league base ball team, or a president of a great banking concern, or a big wheat farmer, or a cattle raiser on a large scale, or the mayor of a great city, or the promoter of some great philanthropic enterprise, or any one of a hundred and one other persons of similar influence and prestige has made a success in life, then the best criterion for estimating scholarship and the best basis on which to predict success in life might conceivably be one's ability to "get by on a pony" in the Latin class, or one's ability to excell in managing college athletics for a season, or even the ability to get others to do the work that one is supposed to do himself. So, also, if success in life is interpreted to mean the becoming of a great painter or sculptor, manifestly aptness in art work and in clay modeling is a better basis on which to predict success than either language or mathematics. Indeed it may be said at the present time that the ambitions of the majority of our brainy young men who enter the high school lead them into fields of endeavor in which little of higher mathematics is wanted and almost none of any kind of foreign language, not to mention the *dead* languages.

For this particular school, however, in which one of the chief purposes is to prepare for college entrance examination in colleges which still persist in requiring a certain formal knowledge of language

and mathematics, together with English literature and a few closely allied subjects, the scholastic standing of the pupils must necessarily be based chiefly upon achievement in these subjects. Even so, there were found only a very few cases in which the rank of a given pupil as based on his S. Q. was markedly different from that given him by a consensus of the teachers' judgments as found on the report cards. There is an almost perfect correlation between the rank of the pupils as determined by the S. Q., as calculated by the above formula, and their rank as determined by an averaging of all the reports made by the teachers.

In the course of our investigation the following facts were brought to light, some of which may have a significant bearing upon certain problems arising in connection with the administration of a private school of this kind in a community like New York City.

The period of time chosen for our investigation comprised the years 1913-14 to 1916-17, inclusive. During this period there were 116 pupils who graduated from the high school, and 184 others who advanced as far as into the eighth grade, in line for graduation in one or another of the four classes, but yet who failed for various reasons to complete the course. Of the 116 graduates 40, or 34 per cent, were boys and 76, or 66 per cent, were girls. Of the 116 graduates there were 16 boys and 36 girls who entered "Early," *i. e.*, before the end of the fifth grade. That is to say, 40 per cent of the boys and 47 per cent of the girls who graduated entered "Early." The records for the four years show a gradually decreasing per cent of graduated boys who entered "Early." The figures for the four years are 44, 43, 40 and 29 per cent. Just the reverse is true in the case of the girls, the per cents being 12, 48, 53 and 53 for the four years, respectively. The increase in the per cent of girls entering "Early" exceeds the decrease in the case of the boys, so that when taken together there is an increase in the per cent. of pupils graduating who enter "early." When combined the per cents are 29, 46, 48 and 49 respectively, for the four years.

As noted above there were, during this same period, 184 pupils who advanced into the eighth grade, in line for graduation in one of the four classes, but who failed to complete the course. Of these 65, or 35 per cent, were boys and 119, or 65 per cent, were girls. Of the 184 non-graduates there were 23 boys and 24 girls, who entered "Early." That is to say, 35 per cent of the non-graduate boys

entered "Early," as against 40 per cent of the graduate boys and 20 per cent of the non-graduate girls entered "Early" as against 47 per cent of the graduate girls. From these data we have calculated the following expectation on the basis of 100 entrants and also the requirements in terms of entrance in order to insure an out-put of 100.

Of every 100 boys who enter "Early" 41 will graduate and 59 will not. Of every 100 boys who enter "Late" 36 will graduate and 64 will not.

Of every 100 girls who enter "Early" 60 will graduate and 40 will not. Of every 100 girls who enter "Late" only 30 will graduate, the other 70 will not. That is to say, of every 100 pupils who enter "Early" 52 are boys and 48 are girls, and of every 100 pupils who enter "Late" 32 are boys and 68 are girls.

That is, if we feed in 100 pupils "Early" and 100 "Late" as they usually come to this school, we may expect 84 of them to be boys and 116 to be girls. We may expect 33 of the 84 boys to graduate and 49 of the 116 girls to graduate.

Or, if it is desired to turn out, say, 100 graduate boys each year, 244 "Early" entrants must be fed in, or 278 "Late" entrants. Or, if the classes are such that it is desired to have an equal number of boys entering "Early" and "Late," it will be necessary to provide 129 of each class, or a total of 258 pupils.

If it is desired to turn out 100 graduate girls each year it will be necessary to feed in 167 "Early" entrants, or 333 "Late" entrants. Or, if it is desired to admit an equal number of "Early" and "Late" entrants the number needed of each class will be 111, or 222 altogether.

The average length of time spent in the grades by the 116 graduates was 6.7 semesters and the average length of time spent in the high school was 7.7 semesters. The average length of time spent in the grades by the 184 non-graduates was 4.61 semesters and 3.8 in the high school. The graduated girls spent on an average 1 semester more in the grades and 1.7 semesters more in the high school than the graduated boys.

Of the non-graduates, the boys spent a slightly longer time both in the grades and in the high school than did the girls. This means that as a rule, in this particular school, the boys who graduate either advance through the high school more rapidly than the girls, or there

are more boys who enter higher up in the high school. The probability is that both these factors operate to make the difference shown in the length of time spent. Referring to our previous statement regarding the per cent of pupils entering "Early," the above supposition is shown to be true, for 60 per cent of the boys who graduate enter "Late" and only 53 per cent of the girls who graduate enter "Late."

Coming now to our main problem, we find that *there were no cases presenting any marked change from "Bright" to "Slow," or vice versa.* as they passed from the "Early" period in the elementary school to the high school, in either the graduate or the non-graduate group. The following procedure was carried out, however, which yielded valuable data for general reference and for the study of certain other specific problems, as that of individual differences in school achievement, as based on sex, age, physical condition, the comparison of teachers' marks in the different subjects in the high school and the relative change in marks from year to year, etc.

The information contained on the high school and elementary record cards was tabulated, in convenient form in a note book after the following manner: A separate tabulation was made for each graduating class, boys and girls separate, from the year 1911-12 to 1916-17, although only the last four years of this period were used in the calculations here presented. The tabulation contained, first the name of the pupil, then the date of entrance and the class entered, the time spent in the grades and in the high school, when graduated, *i. e.*, whether in February or June, standing in the grades and in the high school in terms of "Bright," "Average," or "Slow," the number of each kind of mark received, the S. Q., and finally a distribution of thecases in which he received high marks (A's) and low marks (C-'s and D's) in the various subjects. All the subjects offered in the high school curriculum were grouped into the following eight groups, viz.: English, Modern Language, Ancient Language, Mathematics, Science, Arts, Shop, and Physical Training.

The 52 graduates, who entered "Early," were then arranged in the order of their S. Q. from the lowest to the highest. The S. Q.'s for the whole group ranged from .75 to 2.49. This means that the poorest member of the group is about equal to a student who receives an equal number of B's, C's and D's, and no A's, while the best member of the group is about equal to one who receives an equal

number of A's and B's, with no grades lower. The lower quartile (13 pupils) includes those whose S. Q.'s range from .75 to 1.24, *i. e.*, from the poorest, as described above, to a pupil about equal to one who receives an equal number of all four marks. The lower half of the middle 50 per cent (the second quartile from the lowest) includes the 13 pupils whose S. Q.'s range from 1.24 to 1.54. Thus this group ranged from a pupil about equal to one who receives an equal number of all four marks to one who receives an equal number of B's and C's, with no marks higher or lower. The upper half of the middle 50 per cent (the third quartile from the lowest) includes the 13 pupils whose S. Q.'s range from 1.54 to 1.90. That is from a pupil about equal to one who receives an equal number of B's and C's, with no marks higher or lower, to one who receives all B's, or an equal number of A's, B's and C's, with no D's. The upper quartile, classed as "Bright," include the 13 pupils whose S. Q.'s range from 1.90 to 2.49. The range of this group is thus from a pupil about equal to one who receives all B's, or an equal number of A's, B's and C's and no D's, to one who receives an equal number of A's and B's and no mark lower. We have, therefore, 13 "Bright" pupils, 26 "Average" and 13 "Slow" ones.

An examination of the more than 4,000 report cards for the 13 "Bright" and the 13 "Slow" pupils showed a very close correspondence between the rank as determined by the S. Q. and the rank as based upon the consensus of opinion of all the teachers. If the rank of any given pupil, as based upon his S. Q., is discussed with any given teacher, the individual judgment of said teacher will naturally differ from the S. Q., according as that particular pupil happened to be one who was 'good,' or 'poor' in the particular subject taught by that teacher. But if the consensus of opinion of all the teachers is taken as a basis for rating any given pupil, *no case was found in which a pupil, who was characteristically "Slow" in the "Early" grades, changed so as to be characteristically "Bright" in high school, or vice versa.*

The following two reports, one for a typical "Bright" pupil and the other for a typical "Slow" pupil, represent fairly well the characteristic remarks found on the report cards. The Roman numerals represent the grades from the time the pupil entered the school until the completion of the eighth grade and the capitals represent the four years of high school, which are designated as Alpha, Beta, Gamma and Delta, respectively.

(1) Excerpts from record cards of pupil No. 48, who ranks 9 in the uppermost quartile (The "Bright" group) and whose S. Q. is 2.18.

I. Bright—Capable—Active—Keen—etc.
II. Capable—Older—Serious—Thinks clearly—First impression is of brighter mind than he really is—etc.
III. Bright—Responsive—All round satisfactory.—Right in front rank, yet not above his grade—etc.
IV. Not of manual and mechanical type—Future?—Remarkable reasoning power—Great accuracy and care—etc.
V. Splendid in every way, but lacks in ability to execute with hands—(Very naturally this was taken from a report card handed in by the teacher of manual training.)
VI. Normal—A model—Satisfactory—Keen mentally—etc.
VII. Serious—Excellent—etc.
VIII. Similar reports—etc.
A. Emminently satisfactory—etc.
B. Scholarly qualities—Good worker—Fights against fault of 'Scatterbrainedness'—etc.
G. Typical plodder—Average ability—Steady serious boy—etc.
D. Conspicuous for good scholarship—etc.

(2) Excerpts from record cards of pupil No. 6, whose rank is 6 in the lower quartile (The "Slow" group) and whose S. Q. is .96.

V. Unstable—Inattentive—Changeable—Not very strong—Pass—
VI. Shows improvement—Satisfactory—Careful, but slow—etc.
VII. Long absence—Above average in ability, but below in accomplishment—Lives far out—
VIII. Absent much—Makes excuses—Passing—etc.
A. Average ability—Not strong physically—Easily confused—Tries hard when spirit is good—Lazy?—etc.
B. Failing in Latin and Mathematics—Similar weakness in brother—Lacks application—etc.
G. Reached limit in mathematics—Repeats in Latin, which seems good—
D. Barely passing—Good girl, but weak in scholarship—etc.

In interpreting these reports it should be borne in mind that the pupils who graduate from this school, as a rule, are those who are able to meet the minimum requirements in some, if not all, of the so-called 'heavier' subjects, viz.: language, mathematics, English and science. Moreover, it should be remembered that more of the pupils take these subjects than take the 'lighter' subjects, except in the case of physical training, which is required of all, unless excused by the authorities for physical reasons. Thus it is quite possible for a pupil to attain very high rank in scholarship and at the same time

be reported upon as 'Incapable, Not dexterous, or even possessed with faults' by the teachers of the lighter subjects.

On the other hand, it should be kept in mind that these samples were taken from a group of 52 pupils, all of whom entered "Early" and persisted with at least a passing grade, otherwise they could not have graduated. Thus in case of the "Slowest" of the "Slow," remarks, such as 'Passing, Satisfactory, and the like,' would naturally be found.

As indicated above, a tabulation was made of all the cases in which high marks (A's) and low marks (C—'s and D's) were received by the 116 pupils who graduated from the High School during the period studied, viz.: the years 1913-14 to 1916-17, inclusive. The following is a summary of these data:

	1913–14				1914–15				1915–16				1916–17			
Subj.	Boys		Girls		Boys		Girls		Boys		Girls		Boys		Girls	
	H	L	H	L	H	L	H	L	H	L	H	L	H	L	H	L
Eng.	5	1	3	2	4	4	8	2	2	4	5	3	2	1	13	10
M. L.	4	2	5	1	4	7	16	5	4	2	10	4	1	4	10	13
A. L.	3	–	1	4	3	6	6	4	1	5	4	6	–	4	4	19
Math.	3	5	–	7	4	7	3	13	–	10	1	11	1	2	3	22
Sci.	3	2	3	4	3	8	9	4	4	2	6	6	2	1	6	12
Arts	2	–	5	1	3	1	8	3	2	–	6	1	2	–	14	2
Shop	2	–	–	–	1	1	–	1	3	1	–	–	3	–	–	–
P. T.	4	1	1	–	2	–	7	–	3	–	6	–	2	–	19	2
Total	26	11	18	19	24	34	57	32	19	24	38	31	13	12	69	80
For yrs.	H–44		L–30		H–81		L–66		H–57		L–55		H–82		L–92	

Summary for the 4 years.

	Boys		Girls		Boys & Girls	
Subj.	H	L	H	L	H	L
Eng.	13	10	29	17	42	27
M. L.	13	15	41	23	54	38
A. L.	7	15	15	33	22	48
Math.	8	24	7	53	15	77
Sci.	12	13	24	26	36	39
Arts	9	1	33	7	42	8
Shop	9	2	–	1	9	3
P. T.	11	1	33	2	44	3
Total	82	81	182	162	264	243

The above is not a tabulation of the number of *A's and C-'s and D's* received, but a tabulation of the number of *cases*, in which a pupil received either of these marks. For example, in the class of 1913-14, of which there were 9 boys and 8 girls, there were 26 cases in which A's were given to boys and 18 cases for the girls; there were 11 cases in which boys received C—'s or D's and 19 cases for the girls. For the whole class there were thus 44 cases in which high marks were received and 30 cases in which low marks were received. The summary for the four years shows that there were an almost equal number of cases in which both high and low marks were received by the boys. In the case of the girls, the number of cases in which high marks were received is approximately one eighth more than the cases in which low marks were received. Considering the boys and girls together for the four years there are about one eleventh more cases in which high marks were received than in which low marks were received.

By dividing the number of cases in which high marks were received through by the number of cases in which low markswere received. The ratios thus obtained give us a fair picture of the relative difficulty of each subject, arranging the subjects in the order of these ratios from the highest to the lowest, we have the following array:

Physical Training	14.67
Arts (Dom. & Fine)	5.25
Shop (Man. Tr., etc.)	3.00
English & History	1.56
Modern Language	1.42
Science	.92
Ancient Language	.46
Mathematics	.19

Thus there appears an almost perfect gradation all the way from the subject which is almost wholly a matter of ideo-motor coordination to the subject which involves the largest amount of abstract thinking. The only possible exception is that of ancient language, which should, no doubt, have occurred in the array following modern language. Allowing the facts displayed by this table to stand as they are, the question still might be raised as to the consistency of the standards set up by the various departments. For example, it might be asked, why one subject, as physical training, should be given in such a way that there would be 14.67 times as many cases in

which pupils receive high marks as the cases in which low marks are received, and another subject, as mathematics, be given in such a way that only one fifth as many cases occur with high marks as occur with low marks.

There are three possible ways by which the relative per cent of high and low marks may be made practically the same for all subjects taught. (1) By being more liberal in marking in the case of those subjects which show the lower ratios and by being more exacting in the case of those subjects which show the higher ratios in the list given above. (2) By introducing more difficult subject-matter, in the case of subjects showing high ratios and less difficult subject-matter, in the case of the subjects showing the lower ratios. And (3) by adopting a system of ranking the members of each class from the best to the poorest, using all available information as a basis for such ranking, and then distributing the marks according to a curve approximating that of the normal surface of frequency. Either one of these methods would result in an array of ratios, for all the subjects, varying but slightly either way from any ratio that might be chosen as the most desirable ideal.

In our opinion, the thing that should be done is to combine methods (2) and (3) in all cases where there are as many as 50 pupils in a subject (theoretically there should be at least 100), and with due allowance for exceptional cases in which the selection of pupils in the class is other than random.

A SCALE FOR THE INDIVIDUAL MEASUREMENT OF READING ABILITY

Maud A. Merrill

Research Assistant, Minnesota School for Feeble-Minded, Faribault, Minn.

Experiments in the psychology of reading have shown very strikingly the close relationship between the face or form of the printing type and its legibility (3,11). The total character of the words, the breadth and height and geometrical form of letters, the width of line of the letter, the extent of white margin around the letter, the length of line and space between lines—all condition the legibility of the printed page and, ergo, the reading speed and comprehension. But in the literature of educational measurements but little attention has been paid to this important factor in tests of reading ability. The legibility of the printed page conditions so directly the reading speed that it should not be neglected in the construction of any scale for measuring the reading ability particularly of school children.

In the present reading scale, the factor of degree of legibility has been carefully worked out with a view to making the mechanical conditions of the tests as near the optimum as possible.

The need for standardized reading tests which could be used in the examination of public school children for the special classes was apparent to the writer in connection with the mental examination of such children. In the fall of 1917 under the direction of Dr. F. Kuhlmann at the Minnesota School for the Feeble-Minded, the writer undertook a study of the three R's ability of subnormal children (6). In the course of that study, the present reading scale was constructed to meet the needs of the individual examination which the extant group reading tests failed to supply.

Psychological Requirements in Printing of Reading Scale

In the present reading scale an attempt was made to determine for children of the various grades the reading rate and conditions affecting it. Important among these conditions is the child's manner of perceiving words, which depends largely on how he is taught to perceive them. The reading rate of children of the first grade was found to be almost wholly dependent on the manner of training. That method which emphasizes the apprehension of the whole idea

before beginning to read the sentence aloud, while an almost insuperable obstacle to the measurement of oral reading speed in the first grade, yet seems to show a rapid increase in the lower grades with the attainment of an early maximum reading speed. The child in the first grade has not yet acquired that "fore-boding of the coming grammatical scheme," which James says accompanies reading aloud, and hence often hesitates long before he perceives that "once there was a little girl....!" His silent reading is almost always of the motor type; the early reading of children is universally with the lips moving (9), and there is but little difference between reading silently and reading aloud.

Of first importance in the reading rate, especially for children of the upper grades, is the type used. In a study of the legibility of different faces of printing types (11), Roethlein tried out experimentally some fifty faces of ten point type. These investigations showed News Gothic, Cushing Oldstyle, and Century Oldstyle to be the most legible faces where letters were grouped. In so far as legibility alone is a criterion of excellence in type face, News Gothic would be preferable. Other considerations, however, must be taken into account in selecting a suitable type face for reading. News Gothic has the disadvantage of being an unfamiliar face, bold and unusual in design and character and better adapted for display type. By its very unfamiliarity its usefulness as a reading type for children is limited. Its use is rendered further impractical by its being ill-adapted to the twelve, fourteen, and eighteen point sizes necessary for the various grades. Cushing Oldstyle is similarly ill-adapted for use in the fourteen and eighteen point sizes because one of its distinctive features is a certain heaviness of face which is the result of the thickness of the downward stroke of the letter and while it adds to the legibility of the letter in the ten point type, it renders eighteen point practically useless for reading type.

Century Oldstyle remains, on the whole, by far the most desirable from the standpoint of all practical considerations of leglbility. It has the distinct advantage of being a familiar face, undistinguished by any peculiarities of shading or unusual letter forms, such as, for instance, the Cheltenham G. It conforms to the requirements for legibility in form of letter, size of letter, width of line of letter and extent of white margin around the letter. Increased heaviness of face increases legibility only to a certain point, after which legibility

decreases; broadening the letter form increases the legibility up to a certain point and also increased thickness of line until an optimal thickness of line is reached after which legibility decreases (11).

The size of type to be used for the various grades has been discussed by Huey and Shaw (3, 12) but, to my knowledge, no very extended experiments as to the exact size and kind of type best adapted to the various grades, have been made. Huey, following Shaw's studies, suggests for first grade a minimum size of type[1] of 2.6 mm. with a minimum leading of 4.5 mm.; for grades II and III a type minimum of 2.0 mm., leading 4.0 mm.; for grade IV minimum size type should be 1.8 mm., leading 3.6 mm.; and for grades beyond IV, type and leading should be kept well above the minimal adult requirements. The following table shows the various sizes of the Century Oldstyle type, ten, twelve, fourteen, and eighteen point and the standard requirements in so far as they have been determined.

Table I

	10 point		12 point		14 point		18 point	
	Stand.	Cent.	Stand.	Cent.	Stand.	Cent.	Stand.	Cent.
Size of type	1.5	1.75	1.8	2.0	2.0	2.5	2.6	3.0
Thickness of stroke	.25	.25	–	.3	–	.4	–	.5
Space within letters	.3	.75	–	1.0	–	1.25	–	1.5
No. of letters per cc.	6–7	5–6	–	5–6	–	4–5	–	3–4
Leading	2.5	3.0	3.6	4.0	4.0	5.0	4.5	6.5
Length of line	90–	85.0	–	85.0	–	90.0	–	90.0

In every case the size of type is well above the standard minimum requirement. Leading, the space between lines, has been adjusted to the optimal distance, *i. e.* For ten point type 2.5 mm. is the minimum requirement, but 3.0 mm. is preferable. Thickness of vertical strokes, space within letters (between vertical strokes) and number of letters per running cc. have been standardized for ten point type only.

Requirements for length of line have been very carefully worked out by Dearborn and others (1). Lines of moderate length facilitate speed and ease of reading by lessening the amount of work that the eye must perform in moving from one fixation point to the next.

Footnote 1. Size of type in this discussion means height of lower case letter measured from top to bottom of letter without projections.

Dearborn favors lines from 60 to 80 mm. in length with a maximum of 90 mm. And Huey (3) affirms that "a book should not be used whose lines are more than 90 mm.!" The length of line is relatively of little or no importance for first and second grade readers because the child is concerned almost entirely with the recognition of words and their assimilation into his already acquired "meaning consciousness." Which process has not yet become a habit. With the other grades, however, the length of line is very important indeed, not only from the standpoint of increasing reading speed, but also because it is very important that the child should early acquire the proper reading habits even in the accommodation of eye movements to a sort of "rhythmical regularity" (1) which increases both "speed and ease of reading." Accordingly the length of line for grades I and II and III has been made 90 mm. while for grades IV and V and VI has been fixed at 85 mm. which is well within the maximum of 90 mm.

Century Oldstyle has thus approximated as nearly as possible the ideal mechanical requirements of reading type. Under such conditions it was to be expected that the elimination of all impediments of mechanism would in itself facilitate reading speed. The results seem to confirm this conclusion strikingly.

Method of Procedure

In the standardization of the reading scale, preliminary norms were found, and selection of reading passages made from the examination of the first seven grades in the Faribault public schools.[2]

The object of the present investigation was to standardize reading tests which could be used to determine the reading ability of children, especially in connection with the mental examination. The scale consists of three selections for each of the first six grades.[2a] The scale measures reading ability only as far as the sixth grade for two reasons. In the first place "few feeble-minded children of even the highest type ever do more than fifth grade work successfully. . . ." (2) Subnormal children "show an increasing inability to maintain

Footnote 2. The writer is indebted to Supt. John Monroe and the teachers of the Faribault schools for their courtesy and hearty co-operation in these experiments. We are under obligations, also, to Supt. W. B. Thornburgh of the Owatonna public schools for data on tests of second and third grade children.

Footnote 2A. Owing to limitations of space the selections of the scale cannot be printed here. Anyone who wishes to purchase a copy of the scale may apply to the writer at Faribault, Minn.

the standard that their grade of intelligence would lead one to expect." (6) In reading they attain only 75 per cent. of what they are expected to for their mental age. In an examination of special class children of Detroit, Dr. Renshaw found 86 per cent. below third grade level in reading. (10) These considerations indicate that a reading scale extending as far as grade VI amply fulfills the requirements for such measurements.

In the second place and of much greater importance, is the fact that grade VI seems to mark the upper limit of practical value for an oral reading scale. Huey (3) found average adult silent reading speed to be 2.08 words per second faster than oral reading. Pintner and Gilliland (8) found that even for fifth and sixth grades silent reading showed a slight increase in "reading value," while for third and fourth grades the reading value for both silent and oral methods was the same. Above the sixth grade, however, an increasing reading value was assigned to the silent reading.

The present tests are more directly comparable with Starch's results (13, 14, 15) whose method I have used in expressing reading rate in number of words per second and whose group selections I have standardized as individual tests together with other selections of similar degree of difficulty. My tests show a large increase in reading rate from first to second and second to third grade, with a relatively smaller increase from fourth to fifth and from fifth to sixth, while the increase in silent reading rate (Starch tests) is greater in the upper grades. (cf. Table II). The increase in the silent reading rate in the lower grades is probably due to the difference in the procedure; in silent reading the child will skip a word whose meaning is not known and will in marking the place where he stopped reading (which procedure was used to measure reading speed) be influenced by conditions affecting group tests and will therefor be relatively inaccurate. The examiner cannot, of course, control the conditions of a group test as he can those of an individual examination. It is probable that there is very little difference between the silent and oral reading rate of children of the first grade.

The selections for the scale were made empirically on a basis of comparison with the Starch reading selections (12). Preliminary tests were made only within two months of the beginning of the school year or within two months of the end of the school year. (The final norms are all October scores.) Cases for examination were selected

originally according to age and proper placement in the grade. Only cases within two months of a birthday were examined and only those properly placed in the grade, *i. e.* Eight year olds at the end of second grade were called properly placed in grade II; seven year olds at the beginning of second grade were called standard for first grade, etc. From among the children thus selected, only those were chosen whose school standing in reading was average. This standard of selection proved entirely unsatisfactory not only because it eliminated so many children that there were not enough left, who were available, to establish norms but also because it did not accomplish the object of the selection. Many children in the grades who were either over or under age for the grade were yet doing perfectly satisfactory reading for the grade in which they were placed. Children were then selected on the basis of ability to do satisfactory reading for the grade according to the teacher's judgment. This standard proved much more satisfactory. In the preliminary trials at least six passages were used for each grade. The tentative selections were type written and it is interesting to note in this connection that Roethlein (11) found the American Type-writer face to be the least legible of the types of her investigation. From these six passages were selected for each grade three which showed a maximum and fairly uniform improvement from one grade to the next. These selected passages were then printed according to the requirements indicated and given to children selected as before on the basis of their ability to do satisfactory reading for their grade. All of the preliminary tests were given by the writer. The child was, in each case, given a selection to read and told, "I want you to read this just as fast as you can and just as well as you can." The time was taken with a stop watch and he was allowed to read for sixty seconds. Errors were recorded as follows:

1. Mispronunciations, which indicate unfamiliarity with the word.
2. Substitutions, of another word for the word of the text.
3. Omissions, words left out.
4. Ignoring punctuation marks, indicating lack of comprehension of the text.
5. Hesitations of ten seconds. (If the child hesitated over a word, at the end of ten seconds the examiner pronounced the word for him, recorded the hesitation, which is counted as an error,

and then added ten seconds to his reading time thus insuring that the actual reading time for each selection should be sixty seconds.)

At the end of sixty seconds the number of words read was recorded and the reading rate, in terms of number of words read per second.

The error score has been used rather than any method involving the reproduction of the thought of the selection as an index of comprehension because this method seemed to offer the advantage of being more objective. Without entering into a discussion as to the aim of reading, the attainment of which can be measured only by the correct reproduction of the meaning of the text, we have considered the mental age of the child examined the index of his mental capacity. His reading rate, then, and the number of errors that he made on a selection of known difficulty, were taken as the index of his *acquired* reading skill. It is, of course, true that the child's ability to comprehend the selection directly affects his reading rate and the number of errors that he makes. Quantz (9) found that rapid readers were 37 per cent. superior to slow readers in comprehension. The same is more obviously true of children who have not yet acquired that automatic reading facility which enables the adult reader to proceed with a minimum of consciousness.

The norms, as stated, are all based on October scores. First grade passages were read by children just beginning second grade because it was found that during the first few weeks of school they had had a chance to regain whatever they had lost during the summer vacation months and at the beginning of the second grade represented a completed first grade case. When it came to giving the beginning seventh grade children the sixth grade passages which they had just completed, we found that a new factor of selection entered owing to the fact that in the school system where the tests were tried most extensively, the Junior High School plan was followed, thus widening the gap between sixth and seventh grades. There were fewer poor seventh grade children than sixth, as the promotion from grammar school to high school eliminated many of the poorer pupils who had been pushed on from fifth grade but were not promoted farther.

The tests were given by teachers who knew the children, thus eliminating the possible factor of timidity before a stranger which might influence the brief period of the reading test. The teachers were given careful oral and written instructions and had been pre-

viously trained in the Binet procedure. Given in this way, it was possible to complete the individual examinations during the early part of October.

Results. Table II shows the average number of words read per second by grades, when the selections used were type-written and the increase in number of words read per second when the same selections were printed in Century Oldstyle type according to the standard requirements. The increase in reading rate is very significant in view of the fact that it is consistently greater for each grade.

TABLE II

Increase in reading rates.

Grade	Preliminary (type written)	Final (Cent. O. S. type)
I	1.06 words per second	1.08 words per second
II	1.42 " " "	1.51 " " "
III	1.85 " " "	2.05 " " "
IV	2.16 " " "	2.31 " " "
V	2.44 " " "	2.63 " " "
VI	2.73 " " "	2.81 " " "

Table III shows the increase in reading rate from one grade to the next. Smoothing gives a rate of 1.1 words per second for first grade, 1.5 for second grade, 1.9 for third grade, 2.3 for fourth grade, 2.6 for fifth grade and 2.8 for sixth grade. In determining the reading rate for first grade, the examiner was confronted by the difficulty of overcoming the tendency of some of the chidren (due to the manner of training) to read the sentence silently before reading it aloud. This procedure often took for one sentence more time than was allowed for the entire passage. The child was asked to read it aloud as he went along without waiting to read it to himself first and when this was not successful another method was used[3] but it necessitated omitting the results from one school, for the purposes of the present report, because the results obtained with the alternative method were found to be not strictly comparable.

Foot note 3. This alternative method used for first grade only consisted in determining whether or not the child was familiar with the words of the text. The examiner pointed to the word and the child pronounced it. If the word was unfamiliar or a mistake was made, the examiner allowed a pause of ten seconds before pronouncing the word for the child. The reading time for the passage was taken and the reading rate found by determining, as before, the number of words read per second

Table III

Grade	I	II	III	IV	V	VI
Average	1.08	1.51	2.05	2.31	2.63	2.81
Median	1.05	1.48	2.03	2.27	2.64	2.82
Aver. Dev.	.37	.42	.43	.41	.43	.42
Range	.23–2.7	.37–3.1	1.1–3.6	.93–3.9	1.3–4.0	1.6–3.7

Fifty-eight children were selected for the final norms for first grade; the norms for second grade are based on the examination of sixty-eight selected children; for third grade fifty cases were selected; for fourth grade thirty-seven cases were selected; for fifth grade thirty-seven cases were selected; and for sixth grade there were twenty-two selected cases. Though the number of cases is small they were carefully chosen and we have a double check on the results obtained, by a comparison with Starch's[4] silent reading results and with the preliminary reading rates of the present series.

Beyond fourth grade the increase is not so great in oral as in silent reading. The silent reading speed shows an increasing ratio while in oral reading the speed is already approaching the normal speed for adult readers. Huey (3) found the normal oral reading speed for adults to be from 2.2 to 4.7 words per second. The range of oral reading rates for children in the sixth grade is from 1.6 to 3.7 words per second.

Table IV shows the error scores for the six grades and the relative importance of the various sorts of errors. Errors consisting of omitting words in reading and of ignoring punctuation marks seem to be of the least importance. Errors of mispronunciation and substitution occur with greater frequency, and the range of such errors is much greater. The average total error score for each grade (found by taking the total number of errors of all sorts for each selection) shows great irregularity. If the average number of errors for each grade were constant, then the increasing error score would be due merely to the greater number of words read. However, the score increases from 4.40 to 7.33 from third to fourth grade, then decreases to 7.00 in fifth grade and drops again to 4.05 in sixth grade. The

Foot note 4. (13) Starch's silent reading rates are 1.5 words per second for first grade, 1.8 words per second for second grade, 2.1 for third grade, 2.4 for fourth, 2.8 for fifth, and 3.2 for sixth grade.

decrease in errors in grade VI is probably due to the fact that the children beginning seventh grade who took the test were more highly selected in as much as they had been promoted from grammar schcol to junior high school.

Table VI

Error score by grades

Errors	I			II			III			IV			V			VI		
	Aver.	Med.	Range	Aver.	Med.	Range	Aver.	Med.	Range	Ave.	Med.	Range	Aver.	Med.	Range	Aver.	Med.	Range
Mispronunciation	.8	1	0–5	1.5	1	0–8	1.	1	0–8	3.5	3	0–10	2.6	2	0–8	1.6	2	0–6
Substitution	1.1	1	0–6	1.1	1	0–9	1 .3	1	0–5	1.4	1	0–7	1.7	1	0–8	1.3	1	0–8
Omission	.2	0	0–4	.3	0	0–6		0	0–4	.4	0	0–4	.4	0	0–4	.4	0	0–3
Hesitation	1.0	1	0–4	1.1	1	0–7	.7	1	0–4	.9	0	0–7	2.0	1	0–8	.5	0	0–3
Punctuation	.3	0	0–4	.3	0	0–3		0	0–4	1.2	1	0–6	.7	0	0–5	.2	0	0–1
Total	3.64			4.10			4.40			7.33			7.00			4 05		

Table V shows the degree of difficulty of each of the three selections for which the average rate and error score indicate the average grade attainment.

Table V

Reading rate and average number of errors for each selection

Grade Selection	I Errors	Rate	III Errors	Rate	III Errors	Rate	IV Errors	Rate	V Errors	Rate	VI Errors	Rate
(1)	3.10	1.07	3.63	1.51	4.66	1.96	7.79	2.38	6.89	2.64	4.57	2.76
(2)	3.64	1.11	3.96	1.53	4.10	2.04	7.00	2.26	6.84	2.60	2.85	2.84
(3)	4.20	1.05	4.76	1.48	4.44	2.11	7.45	2.30	7.43	2.66	4.70	2.83

In combining the time and error scores in such a way as to give the best increase in score from one grade to the next, several possible combinations have been tried. Giving E (errors) the same value as R (rate) is obviously incorrect as R shows a constant increase where E is variable. E equals ½ R is very little better as the score very shortly shows a decrease. E equals ¼ R seemed to give the best results so we have used a combination where approximately such a value holds. In table VI the formulæ for the various grades are given using average errors.

Table VI

Reading tests scores

Grade I	10 R−E = 7.3
Grade II	10 R−E = 10.9
Grade III	10 R−E = 13.7
Grade IV	10 R−E = 16.4
Grade V	10 R−E = 19.0
Grade VI	10 R−E = 22.3

R = average reading rate for the three passages for each grade.

E = total errors for the three passages for each grade divided by three to find the average number of errors for each passage.

This formula gives more value to the reading rate and at the same time in the event of an individual case scoring a high reading rate and an unusually large number of errors, evaluates the final result accordingly low.

In order to discriminate a reading ability below the first grade level, which is significant chiefly in dealing with the feeble-minded where it is sometimes desirable to indicate the difference between educability, however slight, and that "pitch black darkness of total ignorance" which even the gentle apologist, Crothers, does not find one of the "Honorable Points of Ignorance," we printed three short reading passages, choosing the vocabulary from primers used in beginning first grade. These passages were standardized by the writer taking the January scores for children in the first grade. The child read the words as the examiner pointed to them; errors consisted of unfamiliar words and ten second hesitations as indicated in the alternative procedure for first grade. The average rate as expressed in number of words read per second is .38, the average error score 4.4. Twenty-six selected cases were used in the standardization of these so-called primary passages.

Conclusions. Reading speed and comprehension are conditioned to an important extent by the legibility of the face of printing type used. The present reading scale printed in Century Oldstyle type fulfills the requirements, in so far as the investigations have determined, of the optimal conditions for the various school grades from first to sixth inclusive.

As a measurement of individual oral reading ability, the present scale shows a consistent increase in reading rate for the first six grades. Using the mental age of the child as the index of his mental capacity, the child should be given the reading passage for the corresponding grade, to measure the degree of attainment of the

mechanical processes of reading, the ability to pronounce words correctly, to read the words of the text without substitutions or omissions and without ignoring punctuation marks. To determine the reading capacity of a case with these standardized passages and scores, he should be tested with the different grade passages until the grade passages are reached for which he does not make the average score.

The present scale should be useful particularly in determining the reading ability of individual cases in connection with the mental examination and is intended especially for such use.

REFERENCES

1. DEARBORN, W. F. *The Psychology of Reading.* 1906.

2. DOLL, E. A. *Clinical Studies in Feeble-Mindedness.* 1917.

3. HUEY, E. B. *The Psychology and Pedagogy of Reading.* 1900.

4. JUDD, C. H. *Measuring the Work of the Public Schools.* Cleveland Educational Survey. 1916.

5. MEAD, C. W. *Silent vs. Oral Reading with One Hundred Sixth Grade Children.* Jour. of Ed. Psych. vol. VI. June, 1915.

6. MERRILL, M. A. *The Ability of the Special Class Children in the 'Three R's.'* Ped. Sem. vol. XXV. March, 1918.

7. PINTNER, R. *Inner Speech During Silent Reading.* Psych. Rev. vol. 20. No. 2. March, 1913.

8. PINTNER AND GILLILAND. *Oral and Silent Reading.* Jour. of Ed. Psych. vol. VII. No. 4, April, 1916.

9. QUANTZ, J. O. *Problems in the Psychology of Reading.* Psych. Rev. Mono., II, 1897.

10. RENSHAW, S. *The Measurement of Abilities of Pupils in the Prevocational Classes.* Report of Investigation of Special Classes, Detroit. June, 1918.

11. ROETHLEIN, B. E. *The Relative Legibility of Different Faces of Printing Types.* Amer. Jour. of Psych. vol. XXIII No. I, January, 1912.

12. SHAW, E. R. *School Hygiene.* N. Y., 1901.

13. STARCH, D. *Educational Measurements.* Macmillan, 1916.

14. .... *Measurement of Efficiency in Reading.* Jour. of Ed. Psych. vol. VI No. 1. January, 1916.

15. .... *The Measurement of Efficiency in Reading, Writing, Spelling and English.* Madison, Wis. 1914.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive reseaiches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3 00 pei year of nine monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X OCTOBER NUMBER 8

EDITORIAL

PROBLEMS IN READING

In the construction of reading scales one of the chief objects is to devise some means of indicating the degree of comprehension of the passage read. In the Starch scale, the Brown scale, and some other scales the fidelity of reproduction is taken as the index of comprehension. This, however, depends upon memory, and its reliability is further invalidated by the fact that the labor of writing, especially in the lower grades, is so distracting as to banish the recollection of much that was comprehended. The degree to which writing interferes with recall varies greatly with different individuals and thus makes comparison of scores precarious as an indication of reading ability. It is natural, therefore, that experimenters should endeavor to reduce the burden of writing to a minimum. In the search for reading material whose comprehension can be indicated by a word or a mark the authors of the Kansas Reading Tests went

to the extreme of introducing arithmetical, physical and practical problems. The solution of these problems undoubtedly depends on the comprehension of what is read, but the mental processes involved are so much more elaborate than those brought into play in the grasping of ordinary reading material that the two activities are scarcely comparable. The Thorndike Scale Alpha calls for answers to questions on the paragraph read. This eliminates the memory factor, and approximates more closely the attitude of ordinary reading, but from Thorndike's own analysis of the mistakes in the answers the test seems to be one of the comprehension of the questions as much as comprehension of the passage read.

Many teachers are skeptical of all these measures of comprehension, and assert that the most reliable indication of reading ability is to be found in the way a child reads a passage aloud. The rate, the smoothness, the inflections of the voice, the expression with which the thought is rendered furnish the most definite indication of comprehension. This, of course, demands individual testing, which is much more costly in time than group testing. It also overlooks the fact that the mental attitude in oral reading is probably very different from that in silent reading. But the most serious drawback to this method is the fact that the estimates of quality of comprehension rest entirely upon the unaided subjective judgment of the listener, and we have no guarantee that the evaluation of the same reading will be the same at different times, or that two judges have the same standards. The only scale for oral reading that we have (Dr. Gray's) deliberately ignores the finer points of expression and voice modulation as indices of comprehension because of the difficulty of applying objective measures to them. The difficulty certainly exists, but is it insurmountable? Would it not be possible by pooling the ratings of competent judges to work out a scale of expression-comprehension in oral reading that could be used to steady the judgments of the individual teacher as do the composition and handwriting scales? In view of the fact that school reading is and will probably continue to be, chiefly oral reading, such a scale would be of great significance in the rating of pupils. If we had a reliable measure of comprehension in oral reading, we could then check up on the mental attitudes in oral and silent reading, and thus come to some more definite conclusions as to the relative values of each.

Again, much of the time now spent in oral reading in schools is practically wasted. Neither pupils nor teacher know just what they are trying to do. The use of such a scale would enable the teacher to diagnose the oral reading needs of pupils more exactly, and thus point the way to more effective remedial measures. Heretofore tests of reading have been considered almost altogether in the mass. There is great need of more intensive psychological studies to bring out the fundamental differences in reading ability, and to determine the degree to which the various scales reveal these differences.

J. C. BELL

NOTES AND NEWS

Superintendent O. S. Hubbard, of Fresno, California, has been appointed director of the Bureau of Research of the Oakland School Department.

Captain Lawrence J. Cole has received his discharge from the army and has returned to his work as professor of psychology in the University of Colorado.

At the New York University Summer School Dr. J. Carleton Bell, of the Brooklyn Training School for Teachers, is giving courses in educational psychology and in educational measurements.

Dr. Daniel B. Leary, formerly head of the department of education at Tulane University, has been appointed professor of psychology at the University of Buffalo.

Mr. H. W. Anderson, statistician of the school system of Omaha, Nebraska, has been appointed to a similar position in Detroit.

Dr. W. Franklin Jones, of the University of South Dakota, has been appointed dean of the school of education of the University of Southern California at Los Angeles.

A Bureau of Cooperative Educational Research will be organized during the year 1919-1920 by the school of education of the University of California in order to make available to the schools of the state the expert services of members of the education faculty.

Lieutenant J. L. Stenquist, who has been on special duty as psychologist examiner in the Air Service of the United States Army, has resumed his work in the Bureau of Reference, Research and Statistics of the Department of Education, New York City.

Dr. Harry Woodburn Chase, professor of psychology in the University of North Carolina, has been elected president of that institution.

Professor Frank E. Morris has returned to the Connecticut College for Women as professor of psychology and ethics. Professor Morris has been serving in the psychological department of the Sanitary Corps of the Army.

At the annual meeting of the District of Columbia Chapter of the Society of the Sigma XI on March 6 Major R. M. Yerkes, Sanitary Corps, U. S. Army, gave an illustrated lecture on the "Relationship of Army Mental Tests to Education and Vocational Guidance."

It is stated that Professor James R. Angell, director of the psychological laboratory, University of Chicago, has been offered the presidency of the University of Michigan, but that owing to disagreement as to conditions the offer has been declined.

The *Institut de Puericulture de la Faculte de Medicine de Paris* was inaugurated on July 1. The Institute was established by the million francs donated to it by the American Red Cross, supplemented by another million francs contributed by French philanthropists.

It is proposed to establish an institute of commercial and industrial psychology and physiology in England. Among the scientific supporters of the undertaking are Professor C. S. Sherrington and Lieutenant Colonel Charles S. Myers. The secretary is Mr. G. Spiller, 1 Great Tower Street, E. C. 3.

The governor of Pennsylvania has recently signed a bill providing for handicapped children in the public schools. Mental and medical clinics are provided to ascertain the children who are in need of special treatment because of defects of vision, hearing or intelligence It is estimated that 150,000 school children need to be examined, although probably not more than 15,000 could be classed as feeble-minded.

Professor L. L. Thurstone, of the Carnegie Institute of Technology, recently issued a revised form of his tests for the psychological examination of high school seniors and college freshmen. They consist of a single folder containing 168 short problems and involve information, mixed relations, completion, syllogistic reasoning, and the explanation of proverbs and statements. The time limit is 30 minutes. The frequent repetition of the same type of test should make the results very reliable in so far as these tests afford a valid diagnosis of intelligence.

The Sixth Annual Conference on Educational Measurements was held at Indiana University on Friday and Saturday, April 18 and 19. The chief speakers were Professor W. W. Charters, of the University of Illinois, and Dr. Walter S. Monroe, director of the Bureau of Cooperative Research, Indiana University. Professor Charters spoke on "The Diagnosis and Correction of Grammatical Errors," "Scientific Curriculum Construction," "Educational Diagnosis and Corrective Instruction," and "Diagnosis and Correction of language."

Dr. Monroe took as his topics "The Progress and Promotion of Pupils in Certain Indiana Cities," "Next Steps in Educational Measurements," "Plans of the Bureau of Cooperative Research," and "Pupils' Errors in Arithmetic." Dr. Ernest J. Ashbaugh, Director of Educational Service of the University of Iowa, spoke on "Educational Service in Iowa," and "Some Recent Developments in Spelling." Other reports were "Results of Group Testing in the Army," "Mental Defectives in the Rural Schools of an Indiana County" by Hazel Hansford, "A Group Scale of Intelligence for First and Second Grades" by Mrs. Sidney L. Pressey, and "School Surveys by Means of Group Tests of Intelligence" by Sidney L. Pressey.

The Sixth Annual Schoolmen's Week convention, held last spring at the University of Pennsylvania, exceeded all records both for attendance and interesting discussions. Among the most notable features was the report of the Bureau of Educational Measurements of the University, presented by Dr. Harlan Updegraff, director, and Mr. L. A. King, assistant director. The announcement that this Bureau had superintended the giving of 200,000 tests in arithmetic and 70,000 tests in silent reading in the schools of Pennsylvania since September, 1918, made a deep impression upon the visiting schoolmen. Not only was this department of the University's School of Education warmly commended, but a more extensive

use was urged of this system. Of the 200,000 tests made in arithmetic, 100,000 were in the Philadelphia schools. This was the first time that the school authorities of Philadelphia had made such tests, and they naturally turned to the University to have them supervised.

Announcement is made by the Carnegie Institute of Technology of the promotion of Lt. Colonel W. V. Bingham to be dean of the division of applied psychology. This faculty includes the departments of psychology, of vocational education and of personnel administration, with which is affiliated the bureau of salesmanship research and the research bureau for retail training. This division has been strengthened by the addition of two new men: Major C. S. Yoakum, of the psychological section of the Surgeon General's Department and formerly of the University of Texas, has been appointed professor of applied psychology and will be director of the bureau of salesmanship research. Lieutenant Colonel Edward K. Strong, Jr., has been made professor of vocational education and assumed on April 1st responsibility for the department for the training of vocational teachers. Colonel Walter Dill Scott will continue his connection with the bureau of salesmanship research as associate director, but will devote only part of his time to the work.

Commissioner Burdette G. Lewis has been authorized by the State Board of Control of Institutions and Agencies of the State of New Jersey, to institute a program of psychiatrical and psychological work that is of special interest to educators and criminologists. It is planned immediately to conduct psychiatrical and psychological examinations in all state correctional institutions and ultimately to extend the work to the other institutions and agencies under the jurisdiction of the state board. The work will be under the general supervision of Dr. Henry A. Cotton, medical director of the New Jersey State Hospital, as acting director of the Division of Medicine and Psychiatry of the Commissioner's staff. The immediate direction of the psychological work of the Division will be under the supervision of Mr. Edgar A. Doll, formerly research psychologist at the Training School at Vineland. All entrants at the correctional institutions will be given a psychiatric test and the Army Group Test. It is also planned, as far as possible, to use the clinical procedures and tests developed in the Army for the psychological examination of recruits. The Army Group Test Alpha, has already been applied to nine hundred men in the New Jersey State Prison, and five hundred public school children of Trenton. It is planned that the psychological examinations shall be the basis of classifica-

tion for the educational, vocational, industrial and parole activities of the correctional institutions of the state. The administrative policies of the several institutions are being re-organized on the basis of the psychological results thus far obtained. The work of the Division of Medicine and Psychiatry will be brought in immediate contact with the Division of Education and Parole and the Division of Labor and Agriculture, of which Mr. Calvin Derrick and Mr. David I. Kelly are the respective directors. Dr. J. M. McCallie, assistant director of education and parole, who has made some extensive researches in the field of applied psychology, will assist in co-ordinating the educational work of the institutions with the educational work of the public schools through the use of these tests. It is planned to extend the plan of work as rapidly as possible. Results thus far obtained are very promising, and indicate that there is an important relation between psychological diagnosis and classification, and institutional management and administration.

PUBLICATIONS RECEIVED

OTIS W. CALDWELL. *The Gary Public Schools: Science Teaching.* New York; General Education Board, 1919. Pp. xix, 125.

This number of the Gary Survey is particularly interesting and valuable to students of the science of education in that almost half of the volume is devoted to the description and discussion of tests in elementary science. So far as the science work in the Gary schools is concerned tha author finds the time allotment fairly generous and the intentions good, but the work is of uneven excellence and suffers from lack of continuity and organizing purpose. Of the nine tests described four were designed to test observation and discrimination, three to test recall and use of previous experiences, and two to test ability to see what occurs and to interpret or give reasons for the occurrences. Test I involved the comparison of a white oak leaf with an elm leaf; test 2, a white oak leaf with a red oak leaf; test 4, the picture of a kingbird with the picture of a kingfisher; test 5, the picture of a kingbird with the picture of a flycatcher. In each of these pairs of comparisons the effort was made to get first a pair that was strikingly different and then a pair that had many points of resemblance. In test 3, the pictures of eight common birds were presented and the pupils were asked to write the name of each. In test 6 ten common birds were named and the pupils were asked to select any four and to tell about each in what kinds of places it builds its nest, of what material and at what time of year it builds its nest, and upon what it feeds its young. In test 7 the names of ten common insects were given, the pupils were asked to select any four, and to indicate about each the kind of place in which it lives, the appearance of the young, the food upon which the young lives, and the harm or good done by the insect. In test 8 the "smoke-box" experiment, a candle placed under an opening in the top of the box caused a current of air which drew part of the smoke from lighted punk through the box. The pupils were asked to tell what was done, what movements of smoke occurred, and what made part of the smoke move through the box. Test 9 showed a funnel passing through a cork fitted into the neck of a tall jar. As water was poured into the funnel the pressure of the air inside the jar tended to balance the weight of water in the funnel. When the cork was loosened slightly bubbles of air escaped and water ran from the funnel into the jar until the equilibrium was again established. The pupils were asked to explain why the water flowed from the funnel into the jar, why it ceased to flow when the cork was pressed into the neck of the jar, and why it flowed again when the cork was loosened. Elaborate directions are given for marking and scoring each test, and the results from various classes are offered as norms. These are the most carefully standardized tests in nature study yet produced.

KATHERINE DUNLAP CATHER. *Educating by Story-Telling.* Yonkers-on-Hudson; World Book Company, 1918. Pp. xx, 396. $1.60.

The story has always been one of the most effective means of imparting information and stimulatin6 the reflective imagination. In this age of dependence upon

books the story is apt to be overlooked and neglected for more direct forms of instruction. This charming book stresses the purpose and aim of story telling, discusses four periods of story interests, describes the building and the telling of the story, and shows how stories may be used to develop an appreciation of literature, music, painting, drama, the Bible, and ethical conduct. Part II contains between thirty and forty stories for telling. The bibliographies are voluminous and well selected, and the book is a treasure-trove for the story teller, whether teacher or parent.

MARION COLLINS. *Case Studies in Mental Defect.* New York State Board of Charities, Bureau of Analysis and Investigation, Eugenics and Social Welfare Bulletin No. xiv, 1918. Pp. 177.

Of the various causes for feeble-mindedness advanced by Tredgold and others psychopathic heredity seems to be the most constant and definite antecedent brought out in these studies. Unusua opportunities were afforded the investigator to observe the evil effects of maladjustment upon individuals and families, and to trace typical anti-social traits in members of the family groups through successive generations. The monograph contains an interesting consideration of the stability of the moron, and an extensive discussion of the sociological significance of mental defect.

ALBERT S. COOK, LIDA TALL AND ISOBEL DAVIDSON. *Course of Study, Baltimore County, Maryland, Public Schools, Grades I to VIII.* Baltimore: Warwick and York, 1919. Pp. viii, 698.

A course of study may serve two functions: A guide to the teachers in that system in which it has been adopted, and a stimulas and source of suggestion to teachers in other systems. One of the greatest needs in education today is to have courses of study made accessible so that they may be compared with each other and the various methods of procedure analyzed. The Baltimore County course of study is noteworthy in several respects. It is sufficiently detailed to indicate not only the material covered in each subject for each grade but also the methods by which the results may be obtained. Each section has exceptionally complete references to the literature of the subject. It is strictly up to date in utilizing the results of the scientific studies of education in class room procedure, as shown in the use of the Thorndike, Ayres and Freeman material in handwriting, the Buckingham and Ayres lists in spelling, and the Courtis tests in arithmetic. The space devoted to history alone is over 150 pages. We predict a wide use of the book for both the purposes mentioned above.

MELVILLE THURSTON COOK. *Applied Economic Botany.* Philadelphia: J. B. Lippincott Company, 1919. Pp. xviii, 261. $1.60.

This text, which according to the sub-title is "based upon actual agricultural and gardening projects," presents a selection of those aspects of botany in which

the general public is most directly interested. The discussion is clear and comprehensible, the text is well printed and sub-divided for study purposes, each chapter is followed by numerous exercises and questions, and the pages are profusely and splendidly illustrated. The book may be used to advantage with either high school or college classes.

George Van Ness Dearborn. *The Psychology of Clothing.* Psychological Monographs, Vol. 26: No. 1, Whole No. 112, 1918. Pp. vii, 72.

Whatever may be the amount of truth in the old proverb that 'clothes make the man' it is undeniable that clothes have a marked psychological effect both upon the wearer and upon the observer. A careful analysis of this effect from the critical point of view of modern psychology is here attempted for the first time. It is a pioneer work, aptly characterized by the author as "a kind of scientific ghost of 'Sartor Resartus' ," and distinguished by the author's well known facility of expression. The subject is treated first as a chapter in physiologic psychology, in which clothing is considered in relation to the skin, to bodily action and to bodily temperature. It is then regarded as a phase of applied psychology and stress is laid upon freedon from fear, the subconscious effects of clothing, self-confidence and success, fitness and the significance of well-fitting clothing, clothes and personality, and the influence of clothing upon initiative.

A. E. Dobb *Education and Social Movements 1700-1850.* New York: Longmans, Green & Co., 1919. Pp. xiv. 257. $3.50.

This is a study of the influences that have affected the development of educational thinking and practice in modern England. The remark that "progress in English education has owed less to the zeal of its advocates than to changes in the structure of social life which have often no apparent connection with educational movements" is probably applicable to other countries as well, and is becoming increasingly applicable as education grows more social in its aim. The author paints a vivid picture of social England at the eve of the industrial revolution, considers the effect of the development of cheap literature upon education, and notes the religious, political and industrial unrest at the close of the eighteenth century. In Part II he gives a sketch of the rise of elementary education in the first half of the nineteenth century, the monitorial system, the mechanics' institute movement, working men's colleges and clubs, the communistic movement of Robert Owen, the growth of libraries, the work of the Tract Societies, the Chartist press, the rise of the democratic ideal, the Hampden Clubs, the increase of manufactures in the nineteenth century, and problems of adolescent and adult education. The book closes with a consideration of modern criticisms of organized education in the light of earlier history. It furnishes a valuable background for understanding present tendencies in English Education.

George Dumas et Henri Aime. *Nevroses et Psychoses de Guerre chez les Austro-Allemands.* Paris Felix Alcan, 1918. Pp. 243. 6 francs.

However violent may be the feelings of the French for the Germans at the present time bitterness does not blind them to the desirability of taking advantage of any observations and studies that the Germans may make in the advancement of scientific knowledge. In the present book the authors present the substance of three reports made by Dr. A. Birnbaum, of Berlin, summarizing the various studies that have been published in technical German periodicals dealing with the normal and pathological psychology of the German peoples under war conditions. These reports were published in 1915-1916, and therefore give us the German reaction to only the first two years of the war. The book opens with chapters on general psychology and psychopathology, discusses war neurasthenia, and devotes the most of the space to a consideration of such psychogenic disturbances as fear neuroses, traumatic neuroses and types of hvsteria.

Fourth Conference on Educational Measurements, Indiana University. Bulletin of the Extension Division, Vo. III, No. 8. April 20 and 21, 1917. University Book Store, Bloomington, Ind., 1917. Pp. 135. $0.50.

The special speakers at this conference were Professor G. D. Strayer, of Teachers College, and Professor C. H. Judd, of Chicago University. Professor Strayer spoke on "The Scoring of School Buildings," "Standardizing the School Plant," "Significance and Present Status of the Survey Movement," and "Practical Improvement in General School Administration Resulting from the School Survey." Professor Judd gave a "Demonstration of the Gray Reading Tests, with Lantern Slides," discussed "Reading Tests," held a round table conference on reading, and spoke on "Experiment in Education," and "Practical Results Obtained Through the Use of Standardized Tests in School Achievement." Other reports presented were "A Study in Arithmetic (Courtis Tests, Series B) in Indiana Cities," by H. L. Smith; "The Value and Use of a Scale in Handwriting (Ayres)," by W. W. Black; A Study in Reading (Gray Oral and Silent Tests) in Indiana Cities," by Cecile W. White; "Suggestions Pertaining to the Training of Feeble-Minded Boys in the Public Schools," by Calvin P. Stone; "Experimental Work in Indiana Schools," by W. F. Book; and "Determination of Standards in Vocational Education," by R. J. Leonard.

Lee F. Hammer. *The Gary Public Schools. Physical Training and Play.* New York: General Education Board, 1918. Pp. xix, 35.

So much has been made of the opportunities for supervised play in the Gary system that one looks eagerly for the results of the survey in this activity. The report considers facilities afforded for play, the teaching staff, the instruction, and the showing made by the pupils in the Athletic Badge Tests issued by the Playground and Recreation Association of America. The time allotted to play is generous and the facilities in the way of equipment are above the ordinary. The

teachers are well trained, but the classes are entirely too large for any real. systematic instruction. On the tests the Gary pupils are below the average of pupils of like age in New Orleans, Seattle, Buffalo and New York in almost every instance. The reason for this the author finds in the amount of time devoted to "free play," which under the mass conditions prevailing often degenerates into mere 'fooling.' On the other hand the Gary pupils seem to be somewhat superior in endurance, as shown by the scores in the last half of their games.

E. Washburn Hopkins. *The History of Religions.* New York: The Macmillan Company, 1918. $3.00.

To present the history of the chief religions of the world in a single volume is a task which demands marked skill in the judicious selection of materials. The author has felt his limitations of space, and has been obliged to pass over many interesting details in the religious beliefs and practices of various countries. On the other hand the account possesses a continuity and carrying power which is lacking in many of the larger and more diffuse works. In spite of his limitation of space the author interpreted his task in a broad sense and endeavored to present a clear picture of the religious consciousness of the entire human race. Thus we find a concise account of the religions of various African tribes, of the Ainus and Shamans (Mongolians), of the Polynesians, of the North American Indians, of the inhabitants of Mexico, Central and South America, of the Celts, the Slavs and the Teutons. The historic religions of India, China, Japan, Egypt, Babylonia, Persia, Israel, Mohammed, Greece, Rome, and Christendom receive more extended treatment. Throughout, the narrative is based on the results of the latest and most exacting scholarship, and voluminous references are given to the authorities. In no way can one obtain such a clear idea of the development of human standards of conduct as through the study of religious beliefs, and this volume is most cordially recommended to psychologists and educators as the latest and in many respects the best book on the subject.

Henry Gustave Hotz. *First Year Algebra Scales.* New York: Teachers College, Contributions to Education, No. 90, 1918. Pp. 87. Cloth, $1.30. Paper, $0.80.

The student of the science of education rejoices to see the gradual spread of the measurement idea to the field of secondary subjects. Rugg, Monroe and Childs have already devised tests in elementary algebra, and in the present volume we have a series of scales which promise to be more precise and convenient than anything yet presented. The monograph makes a brief reference to previous work in the measurement of results in algebra, presents the scales with tentative standard scores, and gives an account of how the scales were developed. The scales are as follows: 1. Addition and subtraction. 2. Multiplication and Division. 3. Equation and formula. 4. Problems. 5. Graphs. Like the Woody arithmetic scales

these scales are arranged in two series, series A, the shorter form, on which twenty minutes are allowed for each exercise, and series B, the longer form, which contains a richer variety of test materials, and for which forty minutes are necessary. Series A is adapted for general school testing, where time is limited, while series B is better suited for individual diagnosis. Both series cover the same range of difficulty, but series B is the richer and more varied. The author emphasizes the fact that these are power rather than speed tests, and that in the time limits fixed the greater number of the pupils should be able to finish the tests. Tentative standard scores are indicated on the basis of the application of the tests to over three thousand pupils who had studied algebra for three, six or nine months.

Ellsworth Huntington. *World-Power and Evolution.* New Haven: Yale University Press, 1919. Pp. 287. $2.50.

This is one of the most stimulating and significant books of the year. The author follows out the argument laid down in his previous book *Civilization and Climate,* that climatic energy is the most important factor in physical environment and that its distribution determines the distribution of civilization upon the earth's surface. The present volume applies this theory to the conditions arising out of the world war. Business activities vary in extraordinary harmony with health, and health depends upon the weather more than upon any other single factor. This thesis is here applied to the explanation of business cycles in this and foreign countries, to a theory of variability in hybrids and the development of new types among animals and men, and especially to the elucidation of periods in literature and other types of mental activity. The rise and fall of the Roman Empire is taken as an example, and the same method of reasoning is then brought to bear upon the solution of the problem of Turkey. In a final chapter the author points out that no nation in the world has so many people who live under a highly stimulating climate as Germany, and attributes to this fact the wonderful power and endurance she showed in holding out for four years against the entire world. This favorable situation, he thinks, will make Germany's influence felt in strong measure in the coming decades.

Herbert Gardiner Lord. *The Psychology of Courage.* Boston: John W. Luce and Company, 1918. Pp. viii, 164. $1.25.

"All normal men are afraid That is why they can be brave. He who is not afraid cannot be brave. He who is brave must of necessity be afraid. The measure of his courage is the violence of his fear successfully overcome." Courage, then, rests on such an overbalancing of the fear impulse by instinctive or reasoned motives that one's objectives are attained in spite of the fear. This discussion of courage is based upon the mechanistic conception of human behavior, and proceeds from the fundamental physiological mechanisms, through the inborn mechanisms, or instincts, (among which we find fear), to the acquired types of reaction, the simpler

and lower forms of courage, and finally to the higher and more complex types. True courage is thus a matter of training, and the special training necessary to develop different types of courage constitutes the climax of the discussion. It is obvious that this is distinctly an educational book.

JOHN BACH MCMASTER. *A Brief History of the United States.* (Revised and enlarged). Cincinnati: American Book Company, 1918. . p. 466, xxx.

This school history, by the author of the encyclopedic *History of the People of the United States,* is marked by accuracy and scholarship on the one hand and by simplicity and directness of style on the other. The text is richly illustrated, well provided with maps, and brought quite up to date. The part played by America in the World War receives due consideration, and the story concludes with the signing of the armistice.

CHARLES MERCIER. *Crime and Criminals.* New York: Henry Holt and Com pany, 1919. Pp. xvii, 290. $2.50.

The author shows little respect for continental European writers on criminology, and adopts an attitude of 'common sense' in regard to the treatment of criminals. He defines crime as "acts or omissions that are infractions of the law, not as it is, but as I conceive it ought to be." This personal interpretation of the law, and personal reaction to criminal situations is characteristic of the entire discussion. As medical officer of lunatic asylums the author has had a certain amount of first hand contact with criminals, and as a writer on mental abnormality he has achieved a certain reputation. His opinions are set forth vigorously and dogmatically, but in no case does he penetrate deeply into the psychology of crime. A characteristic statement is the following: "The first lesson in morality is the cultivation of *esprit de corps.* It is because *esprit de corps* is so thoroughly inculcated and cultivated in our (English) public schools that the education they afford, contemptible as it is from the intellectual point of view, is so invaluable as a moral training."

JAMES BURT MINER. *Deficiency and Delinquency, an Interpretation of Menta Testing.* Baltimore: Warwick and York, 1918. Pp. xiv, 355. $2.25.

Is the mentally defective predisposed to become a criminal, and are all criminals mentally defective? Extravagant claims have been made as to the percentage of feeble-minded in our jails and penitentiaries, and one might be tempted to believe that deficiency and delinquency were synonymous. In the present volume a thorough and scholarly survey is made of the entire question, the evidence of previous studies is carefully sifted, and a method is worked out for evaluating the limit of feeble-mindedness. The author's own studies on delinquents in the Glen Lake Farm School for Boys, Minnesota, are reported in detail, and an extensive com-

parison is made between school retardation and tests of deficiency among delinquents. Part II of the volume is devoted to a discussion of the theory of the measurement of mental development, and to quantitative definitions of the border-line between intelligence and feeble-mindedness. There is a bibliography of 228 titles.

G. H. PARKER. *The Elementary Nervous System.* Philadelphia: J. B. Lippincott Company, 1919. Pp. 229. $2.50.

The behavioristic conception of psychology brings into greater prominence than ever the structure and functioning of the nervous system. In no other way can human behavior be explained than through the action of the neuro-muscular arc An essential phase in the understanding of this neuro-muscular arc is the study of its development in the evolution of animal life, and it is to the problem of the origin and action of this arc in some of the simpler organisms that this study is directed. The volume is the second of the *Monographs on Experimental Biology* issued under the editorship of Jacques Loeb, T. H. Morgan and W. J. V. Osterhout. The discussion deals chiefly with the behavior of sponges, coelenterates and otenophores. In sponges there is no trace of nerve elements, but there is contractile tissue about the pores and the osculum. This supports the thesis that muscular tissue is primary, and that nervous connection is a later development. Studies of the sphincter pupillae and of certain heart muscles confirm the view that contraction under direct stimulation is the primitive type of motor response, and that neural control is a later stage in evolution. The development of neurones in metridium, the aggregation of these neurones into a nerve-net, and the growth of the nerve-net into a well-defined nervous system constitute the chief steps of the argument. There is a bibliography of ten pages. The book is invaluable to the behaviorist.

T. GILBERT PEARSON. *Tales from Birdland.* New York: Doubleday, Page and Company, 1918. Pp. 237. $1.00.

These delightful bird stories make fascinating reading for both children and grown-ups. One does not have to read far to gain the conviction that the author is an authority on bird life, and it is evident from the start that he can tell a good story. These two facts make the book a most desirable addition to the supplementary reading of children in the grammar grades.

RUDOLF PINTNER. *The Mental Survey.* New York: D. Appleton and Company 1918. Pp. x, 116. $2.00.

When an effort is made to find the number of feeble-minded children in a school, it is desirable to avoid needless testing of perfectly normal children. A series of preliminary group tests saves much time and energy. Such a series of tests is here presented. The application of the mental survey to schools and the evaluation of

school achievement in terms of mentality are the most important aspects of the present study. The tests selected include rote memory (Pyle), digit-smybol substitution (Whipple), symbol-digit substitution (Pyle), word building (Whipple), opposites (Thorndike), cancellation (Whipple), language completion (Trabue), and Courtis arithmetic, Series B. Percentile tables of distribution drawn from tests of between three and four thousand children are given for the first tests, and the mental indices by grades are indicated for six schools. The results of the survey tests were checked by comparison with teachers' extimates, and with findings by the Binet and Yerkes scales. Percentile tables of the results with the Trabue and the Courtis tests are also indicated, and these are compared with the findings from the other tests. The second part of the book is a guide for the mental survey, containing directions for giving the tests, for scoring them, and for evaluating the results.

VOLUME X NOVEMBER 1919 NUMBER 9

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

DERIVATION OF A SCALE TO MEASURE ABILITIES IN SCIENTIFIC THINKING*

JOHN P. HERRING

THE measurement of ability in scientific thinking within defined limits has been undertaken by means of a series of problems which appeared in THIS JOURNAL in December 1918. Statements were there made about purpose, structure, limitations, and value, and the history of the test was included. The present article concerns the technique employed in the development of scales for this and similar tests.

I. A PRINCIPLE AND TWO ASSUMPTIONS WHICH UNDERLIE THIS TYPE OF INVESTIGATION

Underlying the interpretation of mental tests is the principle that mental abilities are distributed as are physical traits. If 1000 men, selected at random, pass by in the order of their height, we shall see

first, a very few very short men,
then a greater number of short men,
then a very great number of men of average height,
then a lesser number of tall men,
then a very few very tall men.

If the same men pass by in the order of their ability to remember words, we shall see

first, a very few with very poor memory,
then a greater number with poor memory,
then a very great number with average memory,
then a lesser number with good memory,
then a very few with very good memory.

*This article is a continuation of the author's study *Measurements of Some Abilities in Scientific Thinking*, which is to be found in THIS JOURNAL for December, 1918, Vol. IX, pp. 535-558. The problems there presented should be kept in mind throughout the present discussion.

If the selection is really random, the distribution will take this normal form. "We see the measures distributed somewhat symmetrically about a single, central measure and decreasing in frequency as we pass from the central measure toward either extreme, slowly at first, then more rapidly, then more slowly." (Thorndike, 1913, page 32). Only in the case of an infinite number of measures randomly selected would this curve correspond perfectly to its formula of derivation. Finite groups show irregularities which tend to decrease as the number of measures increases.

There are two assumptions which underlie mental tests of this sort. First, the distribution of the abilities to be studied is normal; and second, the variability of each large randomly selected group of pupils is approximately the same as that of any other.

II. The Probability Curve and its P. E. as a Unit of Measure*

In order to prove the validity of such a test as this, the properties of the frequency curve must be understood. Between the curve and its base are, e. g., all the pupils who took a certain test in arithmetic, arranged according to the number of problems each solved. The height of the curve at a given point (ordinate) represents the number of pupils who solved the number of problems directly below upon the base. Pupils are represented by equal amounts and per cents of area.

The most usual expression of the measures of individuals and of groups has been the per cent, but decimal expression is not always the best. Per cents state exactly the proportion of each group solving each problem and they indicate exactly the order of difficulty, but they are inadequate to establish the exact amount of difficulty. This is for two reasons.

First, the difference between 50% and 60% is equal to that between 90% and 100% as an expression of the number and of the fraction of pupils lying between, but not as an expression of increments of difficulty. In terms of difficulty, per cents lying near 50% are small, while those lying near 0% and 100% are large. It is very hard, not having any ability to spell, to acquire the first notions and the first spellings, and to pass from just not any spelling ability to

*Note: "The P. E. stands for Probable Error, the traditional, but very misleading, name for the median deviation, which is not especially probable, and not an error at all."—Thorndike, 1913, page 40 note.

10%; while it is easy to pass from 40% to 50% ability; but it is very hard again to pass from 90% to 100% or perfection,—that condition in which no words are ever missed and none need ever be studied. This means that as we pass from the left end of the probability curve to the median we see

first, 9.9% of area (or pupils) corresponding to 2.700 P. E.,
then 10% of area corresponding to .652 P. E.,
then 10% of area corresponding to .470 P. E.,
then 10% of area corresponding to .402 P. E.,
then 10% of area corresponding to .376 P. E.,

and as we pass from the median to the right end we see

first, 10% of area (or pupils) corresponding to .376 P. E.,
then 10% of area corresponding to .402 P. E.,
then 10% of area corresponding to .470 P. E.,
then 10% of area corresponding to .652 P. E.,
then 9.9% of area corresponding to 2.700 P. E.

The percent has often been used to measure mental abilities just as if it were a constant unit. It is a variable, and as such, it can be used safely only when the amounts and effects of its variations have been determined.

Second, for want of a common unit, per cents are not comparable from group to group of pupils, except roughly and with an unknown amount of error.

There is, however, a suitable unit for use instead of per cents,—the P. E., or Probable Error. It represents equal steps of difficulty within a group. It is comparable from group to group whenever it can be shown or properly assumed that the groups are approximately, and that their variabilities are approximately equal.

The P. E. of a normal group of unselected measures is that distance below or above the median of its surface of distribution whose ordinate cuts off just 25% of the measures between itself and the median. 1 P. E. corresponds by definition to just 25% of the measures next the median. The median is the chosen point of reference, the place from which we begin to count P. E.'s in both directions; it is just zero distance from the median, and corresponds to just 0% of the measures. It is customary to place − P. E. at the left to represent less difficulty, and + P. E. at the right to represent more. By definition 50% of the cases lie between 1

P. E. and +1 P. E.; 25% lie below —1 P. E. and 25% lie above +1 P. E. The area about —4 P. E. holds the very few pupils who solve very few problems; that between the 25 percentile and the 75 percentile holds the great number of pupils who solve an average number of problems; and that about +4 P. E. holds the very few who solve very many problems. To include all the measures, the curve and the base must be infinitely extended. At infinity they meet. Equal numbers and per cents of pupils are represented by equal amounts and per cents of area in the frequency surface, and equal increments of difficulty by equal distances along the base.

We have then the beginning of a table of correspondences between the per cents or tenths of per cents of the measures, and the fractional parts of the unit P. E. Such a table is Table I†. It contains not only the information that 25% taken from the median, corresponds to 1 P. E., and 0% to O P. E.; but also that 0.1% corresponds to .004 P. E., 10.4% to .391 P. E., and so on.

We have now a unit of measure and a table, I, for converting per cents into multiples of that unit.

A few definitions of terms and relations follow.

The *median* is that perpendicular to the base which so divides the surface of frequency that 50% of the measures lie upon either side. Or, the *median* is the mid measure of the group, when the measures have been arranged in the order of magnitude. Or, "The *median*, or 50 percentile, or mid measure, is the place on the scale reached by counting half of the measures in the order of their magnitude, or the place on the scale above which and below which are an equal number of measures." (Thorndike, 1913, page 43).

†Table I is not reproduced here. See Trabue: Completion—Test Language Scales, 1916, page 38; or Woody: Measurements of Some Achievements in Arithmetic, 1916, page 37. In this table per cents are in the vertical column at the left, and tenths of per cents in the horizontal line at the top. The table reads

.000 corresponds to .000 P. E.
.001 corresponds to .004 P. E.
.002 corresponds to .007 P. E.
.100 corresponds to .376 P. E.
.105 corresponds to .395 P. E.
.357 corresponds to 1.582 P. E.
.499 corresponds to 4.600 P. E.

"The *25 percentile* is the measure with three times as many measures above as below it." (Ibid.)

"The *75 percentile* is the measure leaving one third as many measures above as below it." (Ibid.)

The P. E. is the distance upon the base between the median and the median deviation. It corresponds to 25% of the measures.

Two P. E. is twice that distance. If taken between the median and an ordinate 2 P. E. from the median on either side, it corresponds to about 41% of the measures.

Three P. E. is three times that distance. If taken between the median and an ordinate 3 P. E. from the median on either side, it corresponds to about 47% of the measures.

Four P. E. if taken between the median and an ordinate 4 P. E. from the median on either side, corresponds to about 49% of the measures.

Four and six tenths P. E. taken similarly, corresponds to 49.9% of the measures.

The process of deriving the scale will involve the following steps:

(a) Finding the distance between the medians of successive grades. This problem is solved by means of Tables II, III, and IV.*

*The Roman numerals employed in naming the tables have significance as follows: The data obtained from administering the Scientific Method Test to about 500 pupils in each of grades III to VIII are tabulated in two ways, Table II and Table V. These two tables are the crude data from which are derived all the other tables, except Table I, including the final scale. Table II shows the number of pupils getting each consecutive score from 0 to 33 problems right. Table V shows the number in each grade solving each problem. Tables II, III, and IV, form one series, and Tables V, VI, VII, and VIII, form the other. Table IX is built out of the data from both series, and X from IX.

Since there are two basal tables of data, there are two denotations of the word median. In the first series, which deals with scores, median means median score or median number of problems right, or number solved by a *median able pupil.* In the second series, which deals with the number of pupils solving each problem, median means problem solved by the median number of pupils, or problem just hard enough for just 50% of the pupils to solve it, or *median hard problem.* If such a problem existed for the sixth grade, for example, it would be found in Table VI near problem 13, and would be represented there by 50%. Problem 13 is itself very nearly a median hard problem for the sixth grade. It is represented by .495. Problems 18 and 19 are nearly median problems for the seventh grade.

TABLE II

THE NUMBER OF PROBLEMS RIGHT

Grades	III	IV	V	VI	VII	VIII
33						
32						
31					2	
30						1
29				1	1	
28						
27				1	2	3
26					2	6
25			2		6	14
24			2		11	14
23		1	1	5	13	13
22		1	6	7	28	27
21			7	17	33	35
20		2	7	16	35	42
19		3	20	23	43	43
18		4	28	27	27	57
17	1	5	24	29	52	51
16		8	29	29	45	41
15	3	14	29	45	46	37
14	6	17	29	42	35	35
13	6	23	46	39	37	21
12	4	31	42	33	22	15
11	10	27	41	26	14	15
10	9	39	29	21	16	10
9	20	38	33	17	5	9
8	10	42	31	21	7	2
7	29	33	31	14	6	2
6	34	38	17	11	3	4
5	26	36	24	8	1	
4	36	21	15	6	4	1
3	36	31	10	4	2	
2	39	33	8	1		
1	43	23	6	3	1	
0	50	12	1	3	1	
Number tested	362	482	518	449	500	498
Median Scores	4.362	8.334	12.310	14.417	17.096	18.105

(b) Finding the location of zero.

(c) Finding the location of each problem with reference to each grade median. This is accomplished by means of Tables V, VI, VII, and VIII.

(d) Referring all the problems to zero, using the average of their distances therefrom in the several grades. This is accomplished in Tables IX and X.

III. Finding the Distance Between the Medians of Successive School Grades

By this is meant finding the distances MIII–IV, MIV–V, MV–VI, MVI–VII, and MVII–VIII,* or, finding how far it is from the median of each grade to that of the next.

Three steps appear:

1 How many pupils of each grade between the two medians?
2 Each number is what per cent of its group?
3 What P. E. distance does each per cent represent?

Table II shows the distribution of scores in each grade. For instance, 50 out of the 362 third grade pupils scored O; 43 scored 1; 39 scored 2; and 36 scored 4. It shows how the pupils behaved with respect to the problems. It shows how able are the different grades as tested by the specific series of difficulties, and indicates roughly the degree of normality of the different distributions, and the amount of distance between the different medians.

How many pupils of the third grade lie between MIII and MIV? In Table II we read that MIII is 4.362 and MIV is 8.334. The group of 36 third grade pupils who scored 4, includes pupils barely able to solve 4 problems and pupils barely less than able to solve 5. The group, then, extends from 4.00 problems right through 4.99+ problems right. The score 4.362, therefore, falls within group 4, and in the lower half of the group, and the score 8.334 falls in the lower half of group 8. Count .362 of the way into the group scoring 4. .362 of 36 is 13.032 pupils. These belong logically below the median, and leave, of the 36 pupils, 22.968 logically above the median. Before we get to the next median we must add 23, 26, 36, 29, and .334 of the group scoring 8, or .334 of 10. or 3.34, making a total of 115.3. In Table III we read that it is

*The following key will be of use:

M	median
MIII	third grade median
O	zero
O-MIII	interval between O and MIII
MIII-IV	interval between MIII and MIV

TABLE III.

THE NUMBER AND THE PER CENT, WITH P. E. VALUES, OF EACH GRADE LYING BETWEEN ITS MEDIAN AND THE MEDIANS OF THE OTHER GRADES

GRADES	III	IV	V	VI	VII	VIII
III		115.3	162.2	173.5	180.1	
	.318	.448	.479	.4975		
		1.346	2.411	3.015	4.179	
IV	134.4		141.6	193.1	225.5	230.4
	.279		.294	.401	.468	.478
	1.140		1.217	1.909	2.746	2.986
V	228.6	136.7		87.1	164.3	188.9
	.441	.264		.168	.317	.365
	2.318	1.067		0.644	1.340	1.636
VI	211.3	167.5	79.3		101.3	130.3
	.471	.373	.177		.226	.290
	2.811	1.692	0.681		0.891	1.196
VII	244.5	229.7	183.2	116.4		49.8
	.489	.459	.366	.233		.100
	3.395	2.579	1.643	0.922		0.376
VIII		241.3	201.3	155.4	52.1	
		.4845	.404	.312	.105	
		3.200	1.935	1.313	.395	

115.3 pupils, or .318 of the third grade, are between MIIIand MIV,—P. E. value (Table I) 1.346. 130.3 pupils, or .290 of the sixth grade are between MVI and MVIII, P. E., 1.196 The upper number in each group of three is in every case the number of pupils of the grade indicated at the left who lie between the median of that grade and the median of the grade indicated at the top of the column. The second number is the corresponding decimal. The third number is the corresponding P. E.

115.3 pupils from MIII to MIV, or .318 of the 362 measures in the third grade, or 1.346 P. E. This P. E. distance is recorded in Table IV, together with MIV–III, and a number of other intermedian intervals, including several between non-consecutive medians, such as MIV–VIII. By subtraction of shorter from longer intervals we obtain indirect measures of all the intervals. There are also two direct measures of each interval in each case, that upon each of the two distributions concerned. The discrepancy between the various measures of the same distance is so small, and consequently the effect of unequal weighting would be so slight, that all the measures were assigned the weight of 1, and the averages found. Such an average for MIII–IV appears in Table IV. All the intermedian

TABLE IV.
THE DETERMINATION OF THE INTERMEDIAN INTERVALS

3–4	1.346
3–5 – 4–5	1.194
3–6 – 4–6	1.106
3–7 – 4–7	1.433
4–3	1.140
5–3 – 5–4	1.251
6–3 – 6–4	1.119
7–3 – 7–4	.816
AVERAGE	1.176

Measured upon the third grade distribution, the distance MIII–IV is (Table III) 1.346 P. E.; upon the fourth, 1.140 P. E. These are two measures of the same distance, of which six others are given in the same column. MIII–V, measured upon the third grade, minus MIV–V, measured upon the fourth, is a measure of the distance MIII–IV. The number of the grade upon which the measure is taken always occurs first.

intervals have been computed similarly. By addition they become a scale outline composed of nothing but grade medians. They are as follows:

III–IV	IV–V	V–VI	VI–VII	VII–VIII
1.176	1.176	.597	.868	.365

IV. Finding the Location of Zero

In order to locate zero we must proceed with the intervals O–MIII and O–MIV very much as we did with the intermedian intervals. How far is O below the third grade median? That is, how far is it from the 50 third grade pupils who scored zero to the pupil who got the middle score of the grade? The middle score is 4.362 problems. How many, what per cent, and what corresponding P. E. distance lie between the 50 zeros and the 4.362 problems? Add together those getting 1, 2, and 3 right, and enough of those getting 4 right to equal .362 of the 36 in that group, or 13.032 pupils. The total, 131.032 pupils, lies between the score O and the score 4.362. So the third grade median is 131.032 pupils, or .362 of the third grade pupils tested. This .362 corresponds in Table I to 1.616 P. E., which is a measure of the distance O–MIII.

The fourth grade median is, by means of the same process, found to be 2.905 P. E. above O. The other grades show so few zero scores that they are disregarded. 2.905 minus 1.176 is 1.729 P. E., a second measure of O–MIII.

TABLE V.

NUMBER IN EACH GRADE SOLVING EACH PROBLEM

Grades	III	IV	V	VI	VII	VIII
Number Tested	362	482	518	449	500	498
Problems						
1	231	381	454	405	481	489
2	139	265	360	334	445	445
3	115	280	393	361	423	411
5	133	277	379	349	400	405
4	85	232	325	312	392	424
6	107	198	292	298	401	407
15	89	201	270	319	363	373
7	31	129	287	285	388	446
9	46	118	271	300	373	392
14	65	160	282	288	344	335
8	21	98	237	259	377	400
12	48	151	223	249	319	339
17	59	148	228	220	290	289
13	30	106	229	222	318	332
11	33	96	182	207	276	329
10	42	96	182	175	291	327
16	44	95	191	189	273	317
21	38	108	167	189	239	237
18	34	88	146	168	263	260
19	16	38	87	107	183	205
22	9	37	69	93	191	190
24	20	51	81	98	140	186
28	37	85	119	90	117	116
20	29	67	74	78	132	157
33	44	101	111	77	73	57
23	18	35	74	88	111	148
25	2	13	40	51	148	140
27	9	29	62	68	107	105
26	11	27	58	71	99	100
29	8	10	48	44	125	130
30	21	38	71	50	79	102
31	17	34	47	55	62	77
32	13	18	35	4	32	22

231 pupils of the third grade solved problem 1; 139 solved problem 2. The order of problems is approximately that of difficulty.

These two measures are averaged, the first being given double weight because the 50 pupils of the third grade who scored zero seem to give a firmer indication of the location of zero ability than the 12 pupils of the fourth grade who scored zero.

1.616 P. E.
1.616 P. E.
1.729 P. E.

1.654 P. E.

The weighted average distance O–MIII is 1.654 P. E.

By addition of the intermedian intervals with the interval O–MIII, we have a scale composed of a zero point and six medians.

Grade	Distance Of Medians Above O
III	1.654 P. E.
IV	2.830 P. E.
V	4.006 P. E.
VI	4.603 P. E.
VII	5.471 P. E.
VIII	5.836 P. E.

The scale has now to be completed by the addition to it of the numbers representing the difficulties of the 33 problems themselves. To do this means the production of the final scale and involves the two remaining steps.

V. Locating each Problem with Reference to each Grade Median

How far above and below each median is each problem? To answer this question is the same thing as to construct out of the data of each grade a separate scale of the 33 problems. Table VIII shows these six scales. To produce Table VIII, Tables V, VI, and VII were developed. Table V shows the number in each grade solving each problem. For instance, 231 pupils of the third grade and 381 of the fourth solved problem 1. This table can be more easily studied if each entry is changed to a decimal. This

TABLE VI.

THOUSANDTHS IN EACH GRADE SOLVING EACH PROBLEM

	3	4	5	6	7	8
1	638	790	876	902	962	982
2	384	550	695	744	890	894

0.638 of the third grade pupils solved Problem 1; 0.790 of the fourth grade, etc. Figures for Problems 3 to 33 are omitted.

has been done and Table VI is the result. Those problems which show a constant increase, from grade to grade, in the per cents of pupils solving them, are the best for the purpose. Those which behave erratically have to be regarded with suspicion. Obviously, if the sixth grade scores lower upon a given problem than the fifth, that problem does not measure progress from fifth to sixth grade ability. In the earlier editions of the Scientific Method Test a number of such problems were discovered and eliminated.

In order to make a scale of problems each with a stated numerical difficulty or value, some reference point must be adopted. For inches this would be zero distance. For mental ability, zero, instead of being known to begin with, has to be established, and hence is not available as a reference point. Nor is there any absolute point. A point there is, however, which is fixed relatively to the distribution of the measures about it. It is the position which would be occupied by median difficulty for the group, or the middle point of difficulty, or simply the median, — that is, in Table VI, .500. The median occurs as follows in the several grades:

Grades	Between Problems
III	1 and 2
IV	2 and 4
V	9 and 8
VI	12 and 13
VII	21 and 18
VIII	21 and 18

The deviation of each problem from this point is the difference between .500 and the thousandths in each grade solving each problem. It is each entry in Table VI minus .500. It is, for example, .638 minus .500, or .138; .790 minus .500, or .290. In this way

TABLE VII.

DIFFERENCES BETWEEN 0.500 AND THE THOUSANDTHS IN EACH GRADE SOLVING EACH PROBLEM

	3	4	5	6	7	8
1	138	290	376	402	462	482
2	– 116	50	195	244	390	394

Values for Problems 3 to 33 have been found similarly.

Table VII was derived from Table VI. It is the decimal expression of the deviation of each problem from each grade median.

Table VIII is derived from Table VII by substituting for each decimal entry its corresponding P. E. value, read from Table I. Table VIII shows a scale for each grade. The problems range along the base of each distribution in somewhat the same order and with somewhat the same relative difficulties or distances from each other.

VI. Referring all the Problems to Zero According to the Average Position of Each

But we need a single scale for use in all grades alike. Table VIII gives difficulties with reference to each grade median. We have now to change the reference point to zero for each grade, and to state the difficulty of each problem with reference to zero difficulty, and for all the grades in general instead of each in particular. Table IX does this. Problem I is relatively easier for the eighth grade than for any other, and harder for the third, while for all the grades taken together it will average somewhere between. In finding just where it does average, its position in all the grades is considered and each given a weight. Problem 32 in grade VIII, determined as it is by 4.5% of the children of that grade, is not so reliable a measure as problem 22, in the same grade, measured as it is by 38.2%, or as 21, by 47.6%. Nor is problem I (98.2%) very reliable. Per cents near 0% and 100% are to be distrusted and to be weighted low. In averaging the suggestions of each grade as to the position of each problem, those entries in Table VIII within I P. E. of the median are given a weight of 2, those distant I P. E. or more, but less than 3 P. E., a weight of I, while those distant 3 P. E. or more, are omitted. Only three measures are omitted. Problem I is -.523 P. E. from MIII. MIII is 1.654 P. E. from O. Then by subtraction, problem I is 1.131 P. E. from O P. E. This is according

TABLE VIII.

Grades / Problems	III	IV	V	VI	VII	VIII
1	–.523	–1.196	–1.713	–1.918	–2.631	–3.111
2	.437	–.187	–.756	–.972	–1.819	–1.851
3	.706	–.303	–1.038	–1.269	–1.512	–1.391
4	1.076	.071	–.480	–.756	–1.165	–1.549
5	.504	–.280	–.913	–1.130	–1.248	–1.324
6	.799	.334	–.239	–.628	–1.259	–1.346
7	2.026	.918	–.201	–.512	–1.125	–1.867
8	2.331	1.238	.160	–.288	–1.019	–1.269
9	1.692	1.024	–.085	–.644	–.982	–1.186
10	1.772	1.253	.567	.414	–.307	–.600
11	1.979	1.253	.567	.145	–.194	–.616
12	1.649	.723	.261	–.205	–.523	–.698
13	2.054	1.145	.216	.019	–.516	–.640
14	1.357	.644	–.164	–.539	–.727	–.665
15	1.019	.311	–.018	–.825	–.891	–.996
16	1.728	1.264	.500	296	–.172	–.519
17	1.456	.748	.224	.037	–.299	–.303
18	1.953	1.340	.855	.476	–.097	–.085
19	2.530	2.093	1.427	.057	.508	.330
20	2.083	1.609	1.582	1.391	.936	.710
21	1.859	1.125	.685	1. .296	.082	.089
22	2.905	2.114	1.69	1.211	.445	.445
23	2.439	2.155	1.582	1.269	1.135	.786
24	2.370	1.851	1.449	1.155	.864	.476
25	3.725	2.857	2.114	1.795	.795	.855
26	2.789	2.357	1.803	1.487	1.259	1.243
27	2.905	2.305	1.742	1.524	1.176	1.191
28	1.884	1.380	1.096	1.248	1.076	1.081
29	2.986	3.015	1.962	1.918	1.000	.945
30	2.331	2.093	1.622	1.811	1.487	1.222
31	2.483	2.177	1.979	1.720	1.713	1.506
32	2.2667	2.631	2.211	3.506	2.257	2.514
33	1.728	1.196	1.176	1.409	1.563	1.780

P. E. Deviations of each problem from each median, or P. E. Equivalents of the dfferences between 0.500 and the Thousandths in each grade solving each problem.

to the third grade distribution. According to the fourth, problem I is 1.634 above O P. E.; the fifth, 2.293; the sixth, 2.685; the seventh, 2.840. These results are recorded in Table IX, which gives the

TABLE IX.

	1	2
	1.131	2.091
	1.131	2.091
	1.634	2.643
	2.293	2.643
	2.685	3.250
	2.840	3.250
		3.631
		3.631
		3.652
		3.985
Averages	1.952	3.087

Table showing the distance above zero of Problems 1 and 2 as indicated by each grade, and also as indicated by all the grades averaged together. Problem 1 is 1.131 P. E. above zero as indicated by the third grade. To this indication double weight is given, as explained in the text. Problem 1 is 1.634 P. E. above zero as indicated by the fourth grade; 2.293 by the fifth, etc. In the eighth grade it is more than 3 P. E. from the median, and is therefore disregarded. The weighted average of the various indications is, for problem 1 1.952 P. E., and for problem 2, 3.087 P. E. Problems 3 to 33 have been evaluated in the same manner. The values appear in Table X, Column B.

weighted average position of each problem above zero. Column B in Table X is these distances or P. E. values arranged in the order of their magnitude. This is the scale. The entries in column C, however, are sufficiently accurate and more convenient.

TABLE X.

A	B	C
1	1.952	20
3	3.060	31
2	3.087	31
5	3.090	31
4	3.541	35
6	3.542	35
7	3.920	39
15	3.939	39
9	4.065	41
14	4.146	41
8	4.254	43
12	4.356	44
17	4.492	45
13	4.567	46
11	4.735	47
10	4.749	47
16	4.752	48
21	4.960	50
18	4.991	50
28	5.361	54
33	5.542	55
19	5.561	56
24	5.658	57
22	5.671	57
20	5.708	57
23	5.770	58
30	5.828	58
27	5.874	59
26	5.890	59
31	5.996	60
29	6.194	62
25	6.303	63
32	6.415	64
		1582 Total

A, problem numbers. B, scale of values of the problems, carried to nearest thousandth. C, scale, carried to nearest tenth of a P. E. and multiplied throughout by 10.

AGE-GRADE DISTRIBUTION, ITS PRINCIPLES AND APPLICATION

ERNEST C. WITHAM

Superintendent of Schools, Southington Ct.

THERE are two chief factors in the consideration of the efficiency of any school system. 1. *Quality of Instruction.* This is evaluated best by means of the various standard tests. 2. *Age and Grade Distribution of the Pupils.* Either one of these factors without the other does not tell the story, because we may have a school system which makes a fine showing on all the tests, and yet the retardation may be very great, and the percentage of over-age pupils (pupils too old for the grade they are in) may be very large. There is surely no virtue in a system that takes 2 to 4 years more than is necessary to do its allotted work. On the other hand, a school system that always promote everyone, and in which the percentage of over-age pupils is very small, but where the pupils' work is of inferior quality, such a system would not be efficient, because the small over-age percentage would mean nothing, except that pupils were being promoted regardless of what they know or can do.

It is necessary first of all to have a clear understanding of the terms used in an investigation of this kind.

With regard to age-grade distribution pupils are divided into three groups: 1. Below Normal, 2. Normal, 3. Above Normal.

Below Normal–these pupils too old for their grade.
Normal—these pupils of right age for their grade.
Above Normal–these pupils young for their grade.

The normal ages for grade 1 are 6-7 years, for grade 2, 7-8 years and so on for the remaining grades. All pupils older than these ages are below normal, and all younger than these ages are above normal. These are the most commonly accepted limits of age-grade distributions.

The present investigation covers a period of six years in a growing school system of from 1500 to 2000 children. The information has been tabulated each year by schools and then assembled using the following form:

AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years													
5 "													
6 "													
7 "													
8 "													
9 "													
10 "													
11 "													
12 "													
13 "													
14 "													
15 "													
16 "													
17 "													
18 "													
19 "													
20 "													
Total													
Below Normal													
Normal													
Above Normal													

The age-grade data is always determined the first month of the school year, and the pupils' ages are always figured from September 1. The heavy serrated lines on the blank separate the three groups, the normal group coming between them and the below normal group, below and the above normal group, above the serrated lines.

AGE GRADE DISTRIBUTION FOR SIX YEARS

1912 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	18												18
5 .	96	5											101
6 .	97	52	8										157
7 .	42	78	40	5									165
8 .	11	38	51	32	4								136
9 .	7	12	27	43	36	1							126
10 .	2	8	22	35	39	22	2	1					131
11 .	1	2	5	15	30	33	31	5	1				123
12 .		2	3	10	20	28	36	23	2				124
13 .		1	1	2	12	15	23	35	24				113
14 .		1		1	6	4	13	13	26				64
15 .					1	4	4	3	11				23
16 .						1			1				2
17 .													
18 .													
19 .													
20 .													
Total	274	199	157	143	148	108	109	80	65				1283
Below Normal	21	26	31	28	39	24	17	3	1				190
Normal	139	116	78	78	69	61	59	48	37				685
Above Normal	114	57	48	37	40	23	33	29	27				408

Total retardation 14.8%

1913 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	14												14
5 .	122	3											125
6 .	97	62	2										161
7 .	44	85	41	6									176
8 .	9	49	62	45	5								170
9 .	3	18	35	57	29	3							145
10 .	1	8	12	40	44	32	1						138
11 .			6	18	31	37	22	3					115
12 .		1	4	12	22	26	40	16	4				125
13 .			2	3	7	28	28	40	13				121
14 .					2	10	12	32	29	2			87
15 .					1	8	1	6	27	13	5	1	62
16 .							1	2	11	9	16	2	41
17 .									1	4	9	9	23
18 .										1	6	9	16
19 .												4	4
20 .													
Total	290	226	164	179	141	144	105	99	85	29	36	25	1523
Below Normal	13	27	24	31	32	46	14	8	12	5	6	4	222
Normal	141	134	97	97	75	63	68	72	56	22	25	18	868
Above Normal	136	65	43	51	34	35	23	19	17	2	5	3	433

No. in Opportunity Class 23 Total retardation 14 6%

1914 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	14												14
5 .	118	3											121
6 .	106	54	5										165
7 .	35	81	44	4									184
8 .	7	55	69	35	7								173
9 .	3	24	41	74	35	4							181
10 .	2	4	23	33	38	34	2	1					137
11 .		1	10	22	35	34	23	3					128
12 .		1	4	7	24	31	37	27	3				135
13 .		1	1	1	9	22	27	42	17	6			126
14 .				2		6	17	32	33	12	2		104
15 .							4	8	27	14	5		56
16 .								2	7	16	10	7	42
17 .								1		3	6	16	25
18 .									1		1	9	11
19 .												2	2
20 .													
Total	285	224	197	178	148	131	110	116	88	61	21	34	1583
Below Normal	12	31	38	32	33	28	21	11	8	3	1	2	220
Normal	141	136	110	107	73	65	64	74	60	30	16	25	900
Above Normal	132	57	49	39	42	38	25	31	20	28	5	7	463

No. in Opportunity Class 23 Total retardation 13.9%

1916 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	23												23
5 .	134	2											136
6 .	105	69	1										175
7 .	33	95	58	2									188
8 .	5	39	79	45	2								170
9 .	1	13	38	65	32	6							155
10 .	3	3	28	50	63	21	6						177
11 .	1		10	18	32	31	24	4					126
12 .			1	10	19	40	40	28	2	1			139
13 .				2	4	19	34	37	26	1			130
14 .			1	1	3	19	21	20	32	13	4		121
15 .					1	1	8	8	11	30	4	2	66
16 .									6	17	15	4	48
17 .										2	14	9	26
18 .											3	4	7
19 .										1	1	1	3
20 .													
Total	305	221	213	193	156	136	130	95	77	65	41	20	1652
Below Normal	10	16	40	31	27	39	27	8	5	3	4	1	212
Normal	138	134	117	115	95	71	74	57	43	47	29	13	933
Above Normal	157	71	56	47	34	26	29	30	28	15	8	6	507

No. in Opportunity Classes 42 Total retardation 13 1%

1916 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	17												17
5 .	120	4											124
6 .	118	61	2										181
7 .	44	76	50	1									171
8 .	10	47	76	46	3								182
9 .	1	8	49	77	39								174
10 .		3	17	39	66	31	5						160
11 .			7	23	37	59	34	4					164
12 .			1	13	18	23	37	21	5				118
13 .				3	10	16	33	42	29	1			134
14 .					3	5	18	19	21	19	2		87
15 .				1	1			6	11	25	8	5	57
16 .								1	5	12	22	7	47
17 .										5	12	13	30
18 .											2	11	13
19 .												3	3
20 .													
Total	310	199	202	203	176	134	127	93	71	68	46	39	1668
Below Normal	11	11	25	40	32	21	18	7	5	5	2	3	180
Normal	162	123	125	116	102	82	70	61	32	37	34	24	968
Above Normal	137	65	52	47	42	31	39	25	34	26	10	12	514

No. in opportunity classes 42. Total retardation 10.8%

1917 AGE AND GRADE CARD

Age Sept. 1.	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Total
Under 5 Years	12												12
5 .	154	2											156
6 .	118	45	3										166
7 .	55	90	52	1									198
8 .	17	53	84	28	4	1							187
9 .	3	21	49	61	43	4	1	3					185
10 .		3	27	57	58	36	10	1					192
11 .			8	39	44	56	35	5					187
12 .			1	21	23	38	62	33	4				182
13 .				2	17	16	42	44	25	3			149
14 .				2	3	8	14	18	42	24	2		113
15 .						3	2	8	22	10	19	1	65
16 .									3	8	18	10	39
17 .								1	1	3	7	16	28
18 .											1	7	8
19 .										1		2	3
20 .													
Total	359	214	224	210	192	165	166	113	97	45	47	36	1870
Below Normal	20	24	36	64	43	27	16	9	4	4	1	2	250
Normal	173	143	133	118	102	94	104	62	64	18	25	23	1059
Above Normal	166	47	55	29	47	41	46	42	29	27	21	11	561

No. in Opportunity Classes 40 Total retardation 13.4%

In 1912 it will be noticed that in the second grade there were pupils all the way from 5 years old to 14 years old. Of course such a condition is bad, but was unavoidable until opportunity classes were organized, which now take care of the worst cases of overageness. In 1917 the span of ages for the second grade is 5 to 10, a much better showing.

The percentages of over-age or below normal pupils has apparently not been large for several years at least, yet there has been some improvement from year to year, as a result of careful study of the situation.

The following tables are taken from the age-grade distribution for the past six years.

Above Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	114	57	48	37	40	23	33	29	27				408	31.8
1913	136	65	43	51	34	35	23	19	17	2	5	3	433	28.4
1914	132	57	49	39	42	38	25	31	20	18	5	7	463	29.2
1915	157	71	56	47	34	26	29	30	28	15	8	6	507	30.7
1916	137	65	52	47	42	31	39	25	34	20	10	12	514	30.9
1917	166	47	55	29	47	41	46	42	29	27	21	11	561	30.

Below Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	21	26	31	28	39	24	17	3	1				190	14.8
1913	13	27	24	31	32	46	14	8	12	5	6	4	222	14.6
1914	12	31	38	32	33	28	21	11	8	3	1	2	220	13.9
1915	10	16	40	31	27	39	27	8	6	3	4	1	212	13.1
1916	11	11	25	40	32	21	18	7	5	5	2	3	180	10.8
1917	20	24	36	64	43	27	16	9	4	4	1	2	250	13.4

Normal

	1	2	3	4	5	6	7	8	9	10	11	12	Total	Per Cent
1912	139	116	78	78	69	61	59	48	37				685	53.4
1913	141	134	97	97	75	63	68	72	56	22	25	18	868	57.
1914	141	136	110	107	73	65	64	74	60	30	15	25	900	56.9
1915	138	134	117	115	95	71	74	57	43	47	29	13	933	56.5
1916	162	123	125	116	102	82	70	61	32	37	34	24	968	58.3
1917	173	143	133	118	102	94	104	62	64	18	25	23	1059	56.6

Two methods of attack have been used in studying the above data. 1. Percent of all in each group by grades. 2. Percent in the three groups by grades.

First Method

In the first method we have 100% below normal, 100% normal and 100% above normal. The percent of all the first group in each

grade has been worked out, that is, 11.1% of all the below normal pupils are in the first grade, 13.7% are in the second grade, etc. The following tables give the percentages for the three groups, derived from the first three tables.

Per Cent of the "Below Normals" in Each Grade

Year	Grades 1	2	3	4	5	6	7	8	9	10	11	12
1912	11.1	13.7	16.3	14.7	20.5	12.6	8.9	1.6	.5			
1913	5.9	12.2	10.8	14.0	14.4	20.7	6.3	3.6	5.5	2.2	2.7	1.8
1914	5.5	14.1	17.3	14.5	15.0	12.7	9.5	5.0	3.6	1.4	.5	.9
1915	4.7	7.5	18.9	14.6	12.7	18.4	12.7	3.8	2.8	1.4	1.9	.5
1916	6.1	6.1	13.9	22.2	17.8	11.7	10.0	3.9	2.8	2.8	1.1	1.7
1917	8.0	9.6	14.4	25.6	17.2	10.8	6.4	3.6	1.6	1.6	.4	.8
Ave.	6.9	10.5	15.3	17.6	16.3	14.5	9.0	3.6	2.8	1.9	1.3	1.1

Per Cent of the "Normals" in Each Grade

Year	Grades 1	2	3	4	5	6	7	8	9	10	11	12
1912	20.2	16.9	11.4	11.4	10.1	8.9	8.6	7.2	5.4			
1913	16.2	15.4	11.2	11.2	8.6	7.3	7.8	8.3	6.5	2.5	2.9	2.1
1914	15.7	15.1	12.2	11.9	8.1	7.2	7.1	8.2	6.7	3.3	1.7	2.7
1915	14.8	14.4	12.5	12.3	10.2	7.6	7.9	6.1	4.6	5.0	3.1	1.4
1916	16.7	12.7	12.9	12.0	10.5	8.5	7.2	6.3	3.3	3.8	3.5	2.5
1917	16.3	13.5	12.6	11.1	9.6	8.9	9.8	5.9	6.0	1.7	2.4	2.2
Ave.	16.6	14.7	12.1	11.6	9.5	8.1	8.1	7.0	5.4	3.2	2.7	2.2

Per Cent of the "Above Normals" in Each Grade

Year	Grades 1	2	3	4	5	6	7	8	9	10	11	12
1912	27.9	14.0	11.8	9.1	9.8	5.6	8.1	7.1	6.6			
1913	31.4	15.0	9.9	11.8	7.9	8.1	5.3	4.4	3.9	.5	1.2	.7
1914	28.5	12.3	10.6	8.4	9.1	8.2	5.4	6.7	4.3	4.0	1.1	1.5
1915	31.0	14.0	11.0	9.3	6.7	5.1	5.7	5.9	5.5	3.0	1.6	1.2
1916	26.7	12.6	10.1	9.1	8.2	6.0	7.6	4.9	6.6	3.9	1.9	2.3
1917	29.4	8.4	9.8	5.2	8.4	7.3	8.2	7.5	5.2	5.0	3.7	1.9
Ave.	29.1	12.7	10.5	8.8	8.3	6.7	6.7	6.1	5.3	3.3	1.9	1.5

Figure 1, is a graph of the below normal group of percentages for the six years. The similarity of the curves for the various years should be noted.

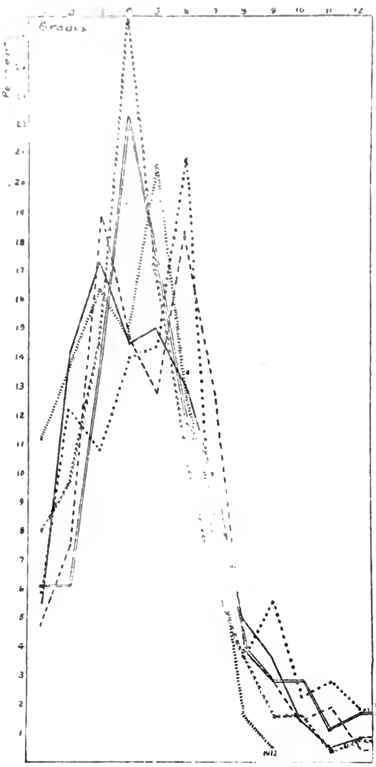


FIGURE 1.

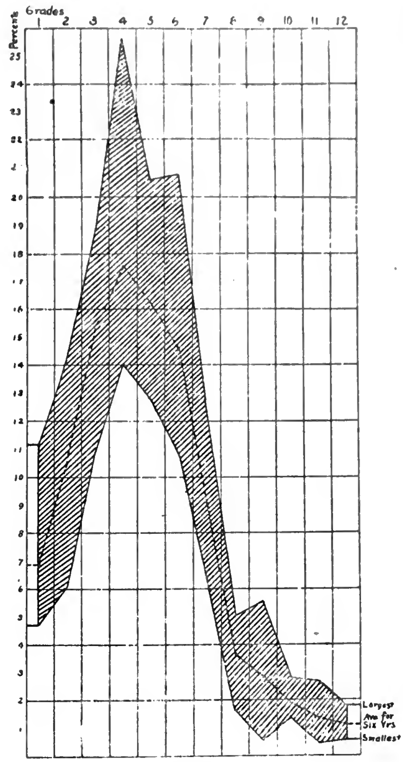


FIGURE 2.

In figure 2 is shown the largest and smallest percentage of below normal pupils for each grade for the six years, and also the dotted line in the center shows the average percentage of below normal pupils for the six year period. This is probably an accurate picture of the condition. It will be noted that the greatest percentages of below normal pupils are in the 3, 4, 5 and 6 grades, but especially in the 4 grade.

What is the reason for so much more retardation in the fourth grade than in any other grade? This is a fair question, and there are probably several answers to it. This is worth thinking about. How do the fourth grades usually come out on the standard tests? All of this is given not to tear down or cast suspicion in any way, but rather in the spirit of constructive criticism, studying the system,

and improving the weakest spots. Is the course of study too difficult for fourth grades? There are several questions that will occur to all concerned at this point. It is our job to solve the problem, the same one that other places are facing if they have yet got to the stage of investigation.

It should be noted that the early age at which pupils are allowed to first enter school, precludes the possibility of much over-ageness in the first two grades.

The school board requires pupils under 16 to finish the sixth grade before leaving school, so that it is easy to see why the intermediate grades have the greatest amount of the below normals. But look at figure 2 and see how many there are in the sixth grade and what a sudden drop there is then. What does this mean? It says very plainly that nearly all pupils who are greatly retarded in the upper grades, leave school just as soon as the law will allow them to. Here are six years of proof that it does not pay to make old pupils repeat the intermediate grades. It simply means driving them out of school at the first opportunity they have for leaving.

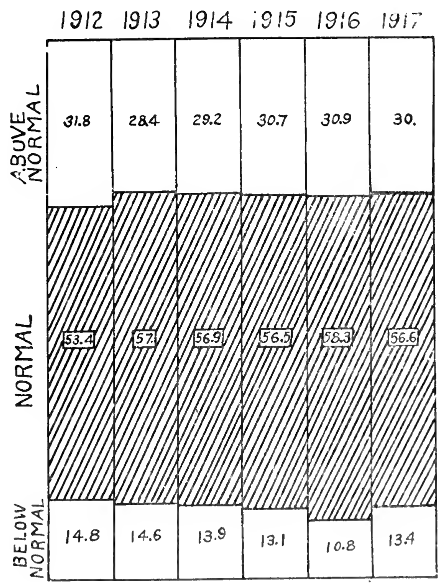


FIGURE 3.

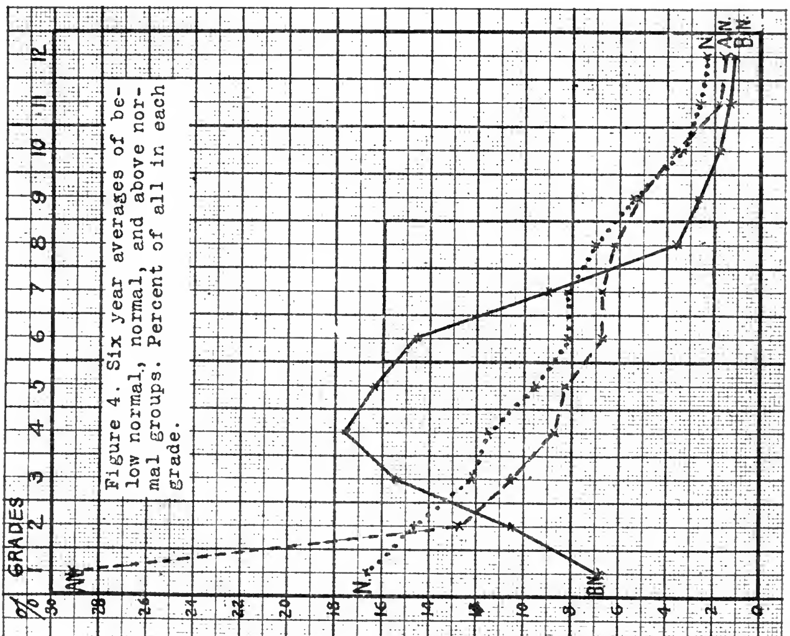


Figure 4. Six year averages of below normal, normal, and above normal groups. Percent of all in each grade.

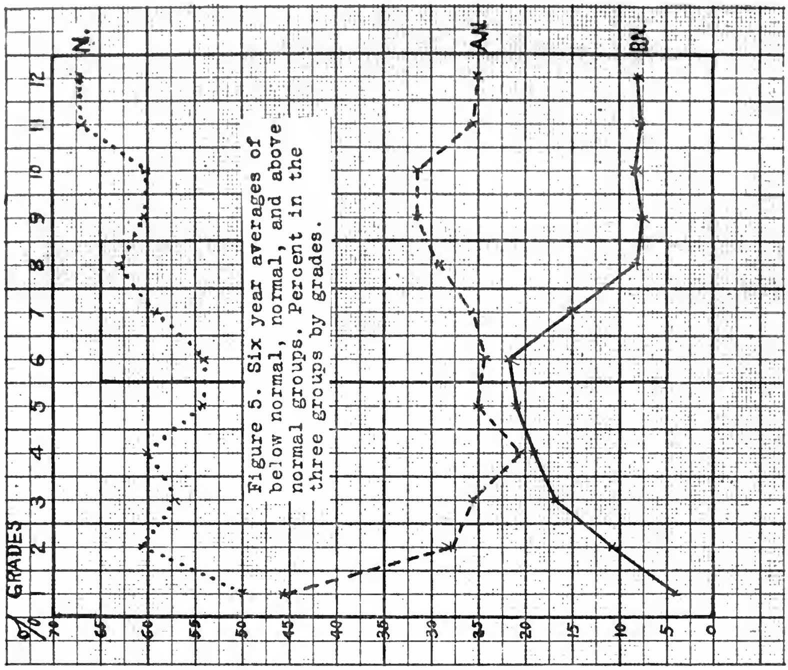


Figure 5. Six year averages of below normal, normal, and above normal groups. Percent in the three groups by grades.

Figure 3 gives both graphically and numerically the percentages of normal, below normal and above normal pupils for the past six years.

Figure 4, gives the graphs of the six year averages of the three groups in percent of all in each group by grades.

The peculiarities of the first grade are to be explained by the following facts:

1. Early admission (If pupils will be 5 during fall term they are permitted to begin school in September).
2. No Kindergartens
3. Large numbers are from non-English speaking homes, so that it takes two years for many to do the first year work.

These facts all tend to make the first grade very large and also to make the above normal group very large.

A comparison of 6-7-8 grade pupils in the three groups show clearly what has already been said, that the Below Normal group leave school while the other two groups of pupils remain in school during these years.

Second Method

The second method used in this study finds out the percent of below normal, normal and above normal pupils in each grade.

Per Cent of pupils in the three groups by grades

Grades	1	2	3	4	5	6	7	8	9	10	11	12
Below Normal	4	11	17	19	21	22	15	8	7	8	7	8
Normal	50	61	57	60	54	54	59	63	61	60	67	67
Above Normal	46	28	26	21	25	24	26	29	32	32	26	25

Figure 5, graphically shows the comparison of three groups. Notice the below normal graph in this figure is very similar to that in figure 4. This shows conclusively that whatever way the below normal data is handled, we get the same general result. Also the comparison of 6-7-8 grades shows even more clearly than in the first method what happens to the three groups.

Comparison with other Places

A large number of school reports have been examined with the view of securing reliable data for comparison with other places. While perhaps 50% of school systems now publish age-grade dis-

tributions, this seems to be done in such a perfunctory way that it is very difficult to find much help in school reports.

In the first place the nomenclature is very different. Some employ the terms Young, Normal and Old; others Accelerated, Normal and Retarded; Under Normal Age, Normal Age and Over Normal Age; and Below Normal, Normal and Above Normal are also used in about the same proportion. Most all use Below Normal to mean the pupils too old for their grade, but there are numerous instances where it is used for the pupils young for their grade.

There is also diversity of practice in the dates employed for securing the data. September 1 and January 1 seem to be the most common now in use.

Practice is about evenly divided in the matter of the arrangements of tabulating the data. Some use the form with the grades horizontally and the ages vertically and others with the opposite arrangements. If we all used the same form we could easily compare results, as it is now comparison is difficult.

Another difference of opinion and practice is in regard to the age limits of the three groups. The majority use 6-7 for the normal ages of the first grade, but 6-6½ and 5-6 and other combinations are frequently found.

Some tabulate age-grade data by straight grades and others by 1a, 1b, 2a and 2b, etc. This would not be so bad but for the fact that in some places an A division stands higher than a B division, while in other places B divisions are higher than A divisions. Sometimes the distribution is made separately for the boys and girls.

After the age-grade data has been secured there is not much evidence that it is analyzed and put to any practical use. Most superintendents seem to stop with just the distribution table. Some few work out the percent of one group, seldom two groups and almost never does anyone work out the percentages of the three groups, but when they do, the second method is the one used. Very seldom is there any comment with the age-grade table. It looks as though it was mechanically assembled because it is in style to have such a table in one's school report, but the facts are apparently not studied, or comparisons made. Even Bliss' excellent recent book on "Methods and Standards for Local School Surveys" does not suggest means of securing age-grade data in a systematic and common way. On page 47 of this book there is a list of about 30 cities giving the

"Number of pupils and Percentage Classification" in age and progress. The facts here presented are of doubtful comparative value because of the differences practiced in procuring the original data.

Information of a reliable nature has been secured for three other cities and is here given.

Percentages of Distribution

	1	2	3	4	5	6	7	8	9	10	11	12	T
1.Hackensack N. J. (Jan. 1917)													
Below Normal	13	13	10	12	12	9	7	14	14				11
Normal	78	78	72	63	54	58	67	65	65				67
Above Normal	9	9	18	25	34	33	26	21	21				22
2. Waterbury Ct. (Sept. 1917)													
Below Normal	8	12	15	20	32	30	17	13	7	7	8	8	18
Normal	88	57	61	56	50	50	62	64	61	55	67	62	60
Above Normal	4	31	24	24	18	20	21	23	32	38	25	30	22
3. Norwich Ct. (Sept. 1915)													
Below Normal	8	9	14	23	27	28	14	13					16
Normal	57	61	64	57	58	54	64	71					59
Above Normal	35	30	22	20	15	18	22	16					25

1. From Annual Report of Public Schools 1917
2. From Annual Report of the Board of Education 1917
3. From Annual Report of Public Schools 1916

The number of cases or records involved in the three cities together with the six years in Southington gives nearly 30,000 cases.

Percent Below Normal

Grades	1	2	3	4	5	6	7	8	9	10	11	12	T
Hackensack	13	13	10	12	12	9	7	14	14				11
Waterbury	8	12	15	20	32	30	17	13	7	7	8	8	18
Norwich	8	9	14	23	27	28	14	13					16
Southington	4	11	17	19	21	22	15	8	7	8	7	8	13
Ave.	8.3	11.3	16	18.5	23	22.3	13.3	12	9.3	7.5	7.5	8	14.5
Percent Normal													
Hackensack	78	78	72	63	54	58	67	65	65				67
Waterbury	88	57	61	56	50	50	62	64	61	55	67	62	60
Norwich	57	61	64	57	58	54	64	71					59
Southington	50	61	57	60	54	54	59	63	61	60	67	67	57
Ave.	68.3	64.3	63.5	59	54	53	63	65.8	62.3	57.5	67	64.5	60.8
Percent Above Normal													
Hackensack	9	9	18	25	34	33	26	21	21				22
Waterbury	4	31	24	24	18	20	21	23	32	38	25	30	22
Norwich	35	30	22	20	15	18	22	16					25
Southington	46	28	26	21	25	24	26	29	32	32	26	25	30
Ave.	23.5	24.5	22.5	22.5	23	23.8	23.8	22.3	28.3	35	25.5	27.5	24.7

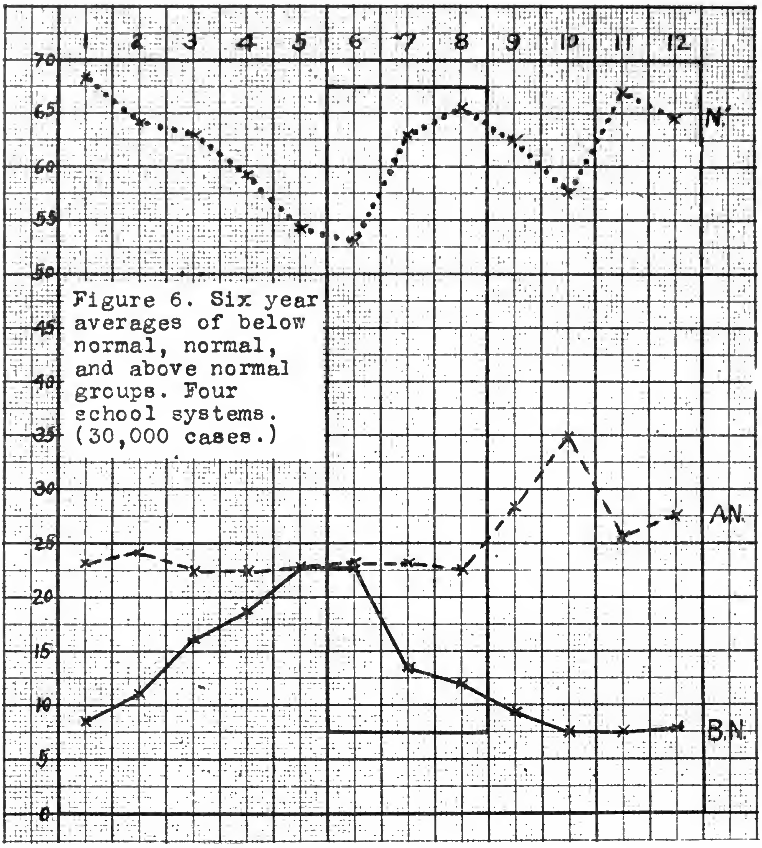


Figure 6. Six year averages of below normal, normal, and above normal groups. Four school systems. (30,000 cases.)

Figure 6, gives the percentage age-grade distribution of the three groups for the four school systems. The below normal graph is very similar to the one in figure 5. Again a comparison of 6-7-8 grades shows that normal and above normal age pupils remain in school while nearly all of the over-age pupils leave school. Some of the latter class return for a few weeks in the seventh grade and so are credited as being in that grade, but they soon nearly all drop out.

VARIABILITY IN MEMORY SPAN*

DAVID MITCHELL

New York City

THE memory span of individuals of the same age has been known to vary in quite marked degree. A person of very low grade intelligence may have a memory span of 3, 2, or 1 or may even be unable to reproduce one impression. On the other hand exceptional individuals have been able to reproduce 10, 11, 12 or even a greater number of discrete impressions. In the same way they vary in every known function. Measurements of ability to read show that in the same age group there are wide variations. These variations are such that in the ten year group there are children who are not able to read any better than some children in the nine year group. Some children of the eight year group excel some of the twelve year group. In the different age groups there is overlapping, the children of one age varying so much that always some of the one age are better than some of those, one, two, three, or four years older.

In a large measure we have tended to ignore this variability and to think that the average performance was typical of every child. The child of eight years should be able to spell certain words, should be able to do certain mathematical operations and should be able to read paragraphs containing words of more or less well known difficulty. All children have been supposed to reach and none have been supposed to greatly exceed this average performance ability. Only recently has this status of fixed averages for age groups been seriously questioned.

The variability in an individual's performance has to a much greater extent been overlooked. Increase in performance ability, corresponding with age increase, has been recognized, but that a child should be better in a performance one day than he would be in a similar performance on the following day has not been generally conceded. If a child is able to read a certain passage on Monday, he should be able to read it as well on Tuesday, even if he has had no practice in the interval, and should read it better if an opportunity for practice has been provided. Furthermore, if he has been able to spell a certain group of words the first day of the month, we

---

*Paper read at The New York Branch of the American Psychological Association, November, 1918.

have assumed that he should be able to spell those words with the same degree of accuracy at the end of the month. If he has been able to repeat correctly the multiplication table on Friday, his ability to do this on the following Monday has not been questioned. Our norms of performance ability have been established on the basis of a belief in the constancy of the individual's performance.

The constancy of average results has confirmed us in our notion of the individual's constancy. The repeated testing of groups of individuals has given average results and measures of deviation which were approximately identical from day to day and from group to group. A study of the individual child and his status from day to day in the group might have shaken our faith in the reliability of individual results, but so far little study of this kind has been made. In order to evaluate the assumption of the constancy of an individual's performance, it was decided to test one function which is supposed to be directly dependent upon neurone excitability. This function cannot be materially changed by practice although it does improve with age. The reproduction of discrete elements, being considered a function of the organic memory, should show as little variation as a function such as mathematical calculation or spelling a certain number of words, since performance in mathematical calculation and in spelling depends largely upon the training or practice which one has had.

Subjects

The children chosen for the experiment were included in five fifth grade groups in one school and three seventh grade groups in other schools. In the five fifth grade classes there were over two hundred children. In the three seventh grade classes there were over one hundred and twenty-five children. Of the fifth grade children, 192 were given a Memory Span test for digits on two days which were separated by a period of exactly five weeks. Of the seventh grade children, 97 were given the Memory Span test on ten successive school days. The fifth grade group included 103 girls and 89 boys; the seventh grade classes 53 boys and 44 girls.

Procedure

The series of digits for each of the trials were made up by a chance selection of numbers, care being taken to eliminate sequences either forward or backward. The presentation of the numbers was given

orally to the group in the various classes and the reproductions were written by the pupils. As the experimenter entered the room at the time of a regular class session, the teacher distributed the record blanks and asked the children to clear their desks and to be certain to have pencils ready. The experimenter then gave instructions in the following way: "We are going to do an experiment this morning. In an experiment we try to find out something. I am going to try to find out how many numbers you can remember. In an experiment it is important that everyone should do the same thing in the same way, so I am going to ask you to pay very close attention to what I tell you and then do exactly as I say. First I want you to write your name at the top of this sheet of paper which has been given to you. Then write your age, your grade, the date, and whether you are a boy or girl." When the children had completed these instructions the experimenter proceeded. "I am going to say some numbers to you and after I have finished saying them I want you to write them down. For instance, if I should say 2-4-6, you would write down at the top of the paper 246 (indicating the method by writing the numbers on the board). Then if I should say 5-7-9, you would write 579 under the other number in this way (indicating them on the board in the right position). You would do this for as many numbers as I might give you. Do you understand? It is important that no one should start to write until I have given all the numbers, so I am going to ask you to hold up your pencils like this (the experimenter holds his pencil up at the level of his shoulder). I will also hold my pencil up and when I bring it down like this (indicating by the movement of the hand to the level of the desk), that will be a signal to you to write down the numbers I have just given you. As soon as you have finished writing them down I want you to hold up your pencil again and that will be a signal to me that you are ready for the next series of numbers. Do you all understand? Then let's begin. Everyone put his pencil up. First I will give you three numbers. Ready!" The test begun, the succeeding series of numbers were given as soon as all the pencils had been raised. The numbers were dictated as distinctly as possible in a monotonous tone of voice at the rate of one per second. The series given were two trials on each series of from 3 to 10 digits.

A method of scoring the Memory Span was a matter of serious question. Consideration was given to the method of scoring suggested by the modified Spearman's Foot rule Correlation $r = 1 - \frac{\Sigma d.}{(n^2-1)/3}$[1]

The amount of labor involved prohibited the use of this method. We therefore, adopted an arbitrary ruling. Two scores were made. The first score was the longest series, in both attempts at which the child was successful at reproduction and in no series previous to which had he failed. This score we may call the "Consecutive Series" score. The other or "Maximum Series" score is the longest series which the child reproduced, regardless of the number on which he had previously failed. A third method of scoring might have been adopted which would give as the child's score the longest series on which the child correctly reproduced one of the trials, and on no series previous to which had he failed on both the trials. Since the question was the variability and not the determination of an absolute standard for different groups the two scores mentioned seemed sufficient.

Results

Tables I and II present the "Consecutive Series" score and the "Maximum Series" score respectively for the 192 children who were tested on two days separated by the period of five weeks.

TABLE I
"Consecutive Series" Score

Trial II / I	3	4	5	6	7	8	9	10	Total
3	2	2	1						5
4	5	16	23	5					49
5	5	20	26	13					64
6	2	6	21	21	10	1		1	62
7			2	6	3				11
8		1							1
9									
10									
Total	14	45	73	45	13	1	1		192

[1]Whipple, G. M. *Manual of Mental and Physical Tests.* Vol. II, p. 161.

TABLE II

"Maximum Series" Score

Trial I \ II	3	4	5	6	7	8	9	10	Total
3									
4		2	2	2					6
5		1	13	18	3		1		36
6		2	17	40	18	8	4	1	90
7			2	7	9	4	4		26
8				1	6	3	1		11
9			1	2		1	1	2	7
10		1	2	3	3	1	2	4	16
Total		6	37	73	39	17	13	7	192

These are correlation tables showing the number of children who made the various scores on the two trials. The scores for trial I are shown in the first column. The scores for trial II are shown in the first line of the table, beginning with the second column. In the table of "Consecutive Series" scores we find that two children had a score of three on each of the two trials; two children scored three on the first trial and four on the second trial; one child whose score was three on the first trial scored five on the second trial. In the third line and third column of Table II we find that two children had a score of four in the "Maximum Series" on the first trial and four on the second trial. The next column of the same line shows that two children scored four and five on the first and second trials respectively. Two children also scored four and six in the first and second trial respectively. In both tables the number of children who had the same score on the two trials is underlined.

The number of children who gave the same score on the two trials, or whose score differed by one, two, or three or more numbers is shown in Table III.

TABLE III

Relation of Scores on Two Trials

"Consecutive Series" Score		"Maximum Series" Score
68	Equal Score	72
100	Different by One	79
20	Different by Two	23
2	Different by Three	9
2	Different by Four	6
0	Different by Five	2
0	Different by Six	1

The first column is for the "Consecutive Series" score. It shows that 68 children had the same score on the two trials. One hundred children scored one more in one trial than they did in the other. Twenty children had scores on the two trials different by two while four children's scores differed by three and four. Similar results are presented in the final column for the "Maximum Series" score. By this method of scoring the variation in the two trials is greater. There were only 151 children whose score was within one of being the same for the two trials. Of the other children 18 varied three or more, two of them varying as much as five and one as much as six.

Table IV shows the record of all children with the averages and medians for both methods of procedure and each trial. The first column of the table gives the scores. The second and third columns contain the numbers of children making the various "Consecutive Series" scores on the two trials.

TABLE IV

Number of Children Making Different Scores

	"Consecutive Series" Score		"Maximum Series" Score	
	Trial		Trial	
Score	1st	2nd	1st	2nd
3	5	14	0	0
4	49	45	6	6
5	64	73	36	37
6	62	45	90	73
7	11	13	26	39
8	1	1	11	17
9	0	0	7	13
10	0	1	16	7
Average	5.1	5.0	6.4	6.5
Median	5	5	6	6

Columns four and five contain the corresponding "Maximum Series" score. The numbers in columns two and three are not identical nor are the numbers in columns four and five. They are sufficiently alike to make the average and the median almost the same for the two trials. The median for the "Consecutive Series" score is the same in both trials and the average differs by only 0.1.

In the "Maximum Series" score the median is again the same, the average score in the two trials differing by only 0.1. The average or group performance for the two days is practically identical. We are not, however, justified in saying that the children show a constancy of performance since they have been shown in Table III to vary considerably.

Tables V and VI show results for the 97 children who were tested on ten successive days.

TABLE V.

"Consecutive Series" Score.

Average Scores and Number of Children Making Them.		Actual Scores on the Different Trials 3	4	5	6	7	8	9	10	0	Total
3.0											
3.5	1	1	6	1	1					1	10
4.0	4	4	24	7	4					1	40
4.5	10	5	36	37	21					1	100
5.0	23	5	52	77	82	12	1			1	230
5.5	26	2	32	48	127	41	8			2	260
6.0	16		13	14	80	40	13				160
6.5	5		4	2	15	19	9	1			50
7.0	9		1	3	20	29	30	7			90
7.5	3		2		3	9	7	6	3		30
8.0											
8.5											
9.0											
9.5											
10.0											
Total		17	170	189	353	150	68	14	3	6	970

In the first part of the first column, the average scores in nearest half units are shown:—that is, if a child made a score of four on six of the ten days, a score of three on two other days, a score of five on the remaining two days, his average score for the ten days would be four. Four children, whose average score individually was four, had scores similar to these. In the second part of the column, the number of children who made the different average scores is shown. In the first line of the table the different scores are shown. For example, in the line for the average score of 7.5, it is shown that three children made this average. Two of the 30 trials gave scores of four, three of six, nine of seven, seven of eight, six of nine, and

three of the trials gave scores of ten. The last line in the column gives the number of times each of the different scores was made in the 970 trials made by these 97 children.

The construction of Table VI is identical, this table being for the "Maximum Series" score.

TABLE VI.

"Maximum Series" Score

Actual Score on the Different Trials

Average Score and Number of Children Making Them.		3	4	5	6	7	8	9	10	0	Total
3.0											
3.5											
4.0											
4.5											
5.0	2		3	8	8	1					20
5.5	1		1	2	7						10
6.0	12		2	14	61	37	6				120
6.5	20			11	84	71	26	7	1		200
7.0	17			2	38	74	43	11	2		170
7.5	17				17	47	81	23	2		170
8.0	9				5	13	37	27	8		90
8.5	9					6	30	37	17		90
9.0	7						15	31	24		70
9.5	3						1	10	19		30
10.0											
Total			6	37	220	249	239	146	73		970

As the average "Maximum Series" score increases, we see that the numbers are sharply cut off in the last column of the table. It was assumed in the beginning of the experiment that very few children would correctly reproduce a series of ten digits. We find that 73 out of the 970 trials result in a "Maximum Series" score of 10. It is obvious from these results that the series should have been increased in length to at least twelve, and possibly a higher number.

The question is raised as to the average daily score in these ten days. Table VII presents the result.

TABLE VII.

Average Daily Scores

Day	"Consecutive Series" Score	"Maximum Series" Score
1	5.5	6.9
2	5.8	7.2
3	4.8	7.3
4	5.6	7.2
5	5.8	7.4
6	6.0	7.6
7	6.0	7.8
8	5.7	7.6
9	5.9	7.6
10	6.0	7.8

In the first column of this table the day on which the trial was given is shown. The second column gives the average "Consective Series" score for the days, and the third column gives the average "Maximum Series" score. In the first case the range of scores is from 4.8 to 6.0, a range of 1.2. This range is wide and should be kept in mind when considering the later discussion of correlations in which the results of trial III are correlated with results from each of the other trials. The reason for this lower score which is 0.8 lower than any other score in the column is found in the fact that the experimenter for one group of children changed his method of procedure for that day. He consciously lengthened the time between the enunciation of the digits in a series. The result was quite marked;—not one of the children with whom he worked succeeded in getting a "Consecutive Series" score higher than four on that day.

The range of variability in averages with the "Maximum Series" score is somewhat less. The lowest score is 6.9 and the highest is 7.8, a range of 0.9. The group results are similar from day to day as was the case with the 192 children whose averages for the "Consecutive Series" score differed by only 0.1 and the averages of the "Maximum Series" score showed only the same difference. Group results are approximately identical from day to day but individual scores vary considerably.

In order to get a simpler indicator of individual variabilities coefficients of correlation were calculated according to the formula

$$r = \frac{p^{2}}{\sigma \times \sigma y}$$

[1]See page 181ff. *Introduction to the Theory of Statistics*, by G. Udny Yule.

TABLE VIII.

Coefficients of Correlation

Trials	"Consecutive Series" Scores	"Maximum Series" Scores
I-II	0.40	0.54
I-III	0.09	0.49
I-IV	0.46	0.50
I--V	0.30	0.59
I-VI	0.19	0.64
I-VII	0.31	0.59
I-VIII	0.31	0.50
I-IX	0.38	0.50
I-X	0.46	0.52
II-III	0.30	0.49
II-IV	0.45	0.61
II-V	0.47	0.71
II-VI	0.42	0.55
II-VII	0.45	0.62
II-VIII	0.27	0.49
II-IX	0.42	0.48
II-X	0.60	0.53
III-IV	0.41	0.51
III-V	0.44	0.57
III-VI	0.46	0.61
III-VII	0.18	0.63
III-VIII	0.17	0.46
III-IX	0.26	0.58
III-X	0.16	0.52
IV-V	0.60	0.60
IV-VI	0.49	0.58
IV-VII	0.37	0.62
IV-VIII	0.41	0.44
IV-IX	0.47	0.67
IV-X	0.38	0.46
V-VI	0.33	0.65
V-VII	0.47	0.73
V-VIII	0.38	0.57
V-IX	0.53	0.62
V-X	0.51	0.51
VI-VII	0.28	0.67
VI-VIII	0.33	0.63
VI-IX	0.39	0.58
VI-X	0.43	0.65
VII VIII	0.29	0.56
VII--IX	0.44	0.63
VII-X	0.54	0.49
VIII-IX	0.48	0.40
VIII-X	0.35	0.57
IX-X	0.64	0.48
Average	0.37	0.54

The correlations for the two trials with the 192 children were 0.47 and 0.44 for the two methods of scoring. For the ten trials with the group of 97 children we have a total of 45 coefficients of correlation. They are given in order in Table VIII.

The first column gives the trials correlated and the second and third columns the coefficients for "Consecutive Series" and "Maximum Series" scores respectively. The values are somewhat less than they should be for one reason and somewhat more than they should be for another. As mentioned before, one of the three experimenters changed his method for the third trial so that not one of forty children scored higher than four on the "Consecutive Series." All these correlations, therefore, in which the third trial is considered are lower than we might expect. On the other hand, our series of digits did not go lower than three. A number of children failed in getting this series correct so that their "Consecutive Series" score could only be zero. Since this break occured in the series it did not seem wise to include the zero score in the correlation. The coefficients are, therefore, somewhat higher than we might expect if series of digits down to one had been used.

A collation of the coefficients is shown in Table IX.

TABLE IX.

Coefficients of Correlation	Coefficients of Correlation "Consecutive Series"	"Maximum Series"
0.00–0.09	1	
0.10–0.19	4	
0.20–0.29	4	
0.30–0.39	12	
0.40–0.49	18	10
0.50–0.59	3	18
0.60–0.69	3	15
0.70–0.79		2
Total	45	45

The first column shows them grouped in ranges of ten points. The second and third columns contain the number of the correlations. According to this table the coefficient of correlation for one pair of trials was less than 0.9. For four trials it ranged between 0.10 and 0.19; for another four between 0.20 and 0.29. In only six cases

is the coefficient of correlation for "Consecutive Series" scores greater than 0.50. The average is 0.37. The average coefficient of correlation for the "Maximum Series" score is 0.54. Thirty-three come between 0.50 and 0.69; only two are greater than 0.70. If the memory span were a constant function, one would be justified in expecting a higher correlation than those that are shown. A coefficient less than 0.75 is extremely significant when different functions are being correlated, but correlations of less than that value of the same function on different days leads us to suspect that the apparent validity of the individual test may be seriously questioned.

Conclusion

In our evaluation o fmemory span we have been largely concerned with group or average results. We have determined what the score for children in a great many different problems should be. We have assumed, if a child reaches that average score in one examination, that he has the ability to do it on all occasions. We, therefore, claim that he ranks in intelligence in a certain part of the intelligence scale. On the other hand if in one examination a child fails to make the average score, we have judged him incapable of doing it and ranked him in a lower part of the intelligence scale. Our results indicate that this method of grading needs revision. The performance ability of children is not a static or fixed quantity. It rather varies from month to month, from day to day, or we might even say from hour to hour. A child who is able to make 100% in Spelling one day may make only 75% or 50% in Spelling the same list of words on another day. Children who are able to do 35, 40 or 50 problems a minute one day may do considerably less or considerably more on another day.

When we consider the question from the standpoint of average performance for a group, we do not need to be concerned with this variability. The average score for a group remains approximately constant with the same degree of practice, or increases in fairly constant ratios with similar amounts of practice. Children in one city with the same amount of practice as children in another city give almost identical results. But the individual child may materially change his own position in the rank order. He who is first may not change his position so that he is last but he may change it to such an extent that neither the individual score nor the average

of two scores will give a satisfactory record of his ability or his intelligence, or even of his capability in the specific function tested.

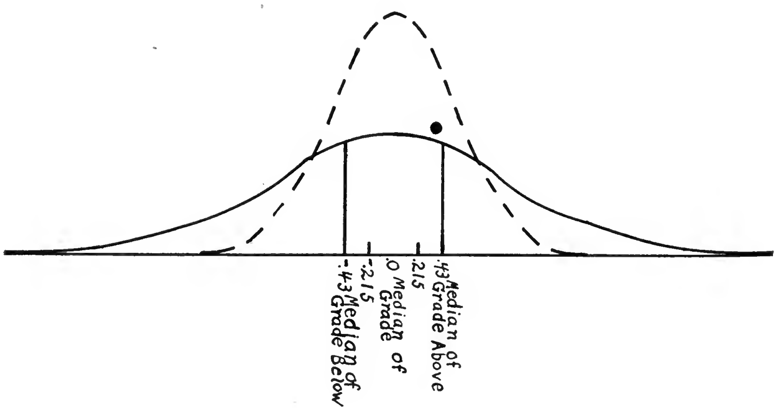
These results may not be considered as conclusive evidence that our mental testing is being done under the handicap of a false assumption. It indicates however, that a revision of our ideas concerning it may be necessary. In the scales of intelligence which have been devised, great weight is placed upon the ability to reproduce numbers and also to reproduce series of words in sentences. The method of giving the test and of scoring probably eliminates some of the variability which we have found, but when the "Maximum Series" score of children can vary as much as six within the range of ten, it does seem that injustice may be done to the child who on a certain day gives his lowest possible score in the memory span. Before the evidence may be considered conclusive, it will be necessary to determine the amount of variability when the method followed is identical with that prescribed in the intelligence scale. There is little doubt, however, that the variability in performance merits careful and serious consideration and demonstrates that the doubt should always be in favor of the child until a wider range of experiment with him has been possible.

COMMUNICATIONS AND DISCUSSIONS

THE MEASUREMENT OF OVERLAPPING*

All who have given standardized tests or have used scales for the measurement of accomplishment of children in different grades have found great variability in each grade and extensive overlapping of grade upon grade. The existence of extreme overlapping in reading ability is reported by Daniel Starch[1] who writes "it is found that in speed and comprehension 31.8 per cent of the pupils of any grade reach or exceed the median of the next grade above," and by F. J. Kelly, who reports[2] that "The reading ability possessed by the median child in any one grade is superior to the ability possessed by at least one third of the children in the class above him and is no better than the reading ability possessed by at least a third in the class below him."

Assuming a normal distribution of scores in reading ability and calling the standard deviation of such scores σ, this statement would imply, as may be determined by reference to a table of the frequencies in a normal distribution, that one grade median is .43 σ distant from the next. The situation is pictured in the accompaning full-line graph.



Let us assume that the score in the reading test is a valid criterion of ability, and determine the per cent of pupils who would be better classified if they were in the grade above or the grade below. Any pupil making a score greater than that which corresponds to .215 σ

*This communication was printed in the April, 1919, issue of the JOURNAL. Owing to an oversight the proof was not read, and the formulae were quite unintelligible. In order to make such amends as are possible, the article is reprinted here. EDITOR.

[1] DANIEL STARCH. *Educational Measurements.* p. 42.

[2] *The Kansas Silent Reading Tests.* J. ED. Psyc. v. 7 No. 2, 1916.

is closer to the median accomplishment of the grade above than to that of the grade in which he is located, and therefore should be transferred to the grade above. The percentage of cases in a normal distribution lying above .215 σ is 41.5, and similarly 41.5 lie below—.215 σ. Accordingly, by this criterion, 17 per cent are properly and 83 per cent improperly classified. As, thruout a number of grades, reading ability is presumably the most important single determinant in classification, so extreme a finding as this may well be questioned and carefully considered before finally accepted.

We may approach this matter from another point of view. If we give a five minute test in solving puzzles we are testing ability to react to novel situations and may call our test a reasoning ability test. From data at hand the writer knows that enormous overlapping would be found with such a test. The reason is here quite obvious. The puzzle ability of an individual may be quite identical with his reasoning ability, but it cannot be determined in five minutes. The probable error of a five minute determination of puzzle ability age is very likely three or four years. The great overlapping in this case is seen to be a consequence of the unreliability of the test, and not the trivial nature of the function tested or of the real overlapping of talent from grade to grade.

To a degree the same situation maintains in the case of every measure of overlapping, unless it be in dealing with such measures as height or weight which can be "exactly" measured, and the unreliability of measures must be known before significant estimates of overlapping can be made.

The writer has found in the case of two classes of about thirty, that the reliability coefficient (the extent to which the test correlates with a second similar test) of both the Starch Reading test and the Kansas Silent Reading test was between .2 and .3. More extensive determinations of these reliability coefficients would probably not change the values appreciably, so it is clear how it has happened that such great overlapping was found.

Knowing the reliability coefficient, it is possible to determine what the overlapping would be if measured by means of a very large number of such tests. We will call this the true overlapping with respect to the ability tested. If the scores of the members of a class are the average scores based upon a very large number of tests, all similar to each other in difficulty and in what each tests, the scores will truly represent ability and whatever overlapping is then found will be overlapping in true ability. Consider the dotted curve shown to represent the distribution of scores based upon averages of a large number of tests. If the grades received in the first test are designated by x_1 (x_2 in the second test, x_3 in the third, x_n in the n'th), where x_1 is a deviation from the group mean, then the full line curve represents the distribution of grades x_1 received

by the members of the class (consider N to be the number of individuals), and the dotted curve represents the distribution of the N average grades, each of which is of this type, $\frac{x_1 + x_2 \cdots + x_n}{n}$. We wish to know the spread or dispersion of these latter measures and will therefore calculate their standard deviation:

$$\sigma^2 \text{ of } \left(\frac{x_1 + x_2 + \cdots x_n}{n}\right) = \frac{\Sigma\,(x_1 + x_2 + \cdots x_n)^2}{nN}$$ in which Σ indicates a summation of N terms.

$$= \frac{\Sigma x_1^2 + \Sigma x_2^2 + \cdots \Sigma x_n^2 + 2\Sigma x_1 x_2 + 2\Sigma x_1 x_3 + \cdots 2\Sigma x_1 x_n + 2\Sigma x_2 x_3 + \cdots}{n\ N}$$

Since according to hypothesis the tests which are averaged are all similar they all have the same standard deviation and correlate with each other to the same amount; that is to say

$\Sigma x_1^2 = \Sigma x_2^2 = \cdots = N\sigma_1^2$, according to the usual definition of standard deviation, and

$\Sigma x_1 x_2 = \Sigma x_1 x_3 = \cdots = N r\, \sigma_1 \sigma_2$, according to the usual definition of a product-moment coefficient of correlation.

In this last equation r is the correlation between one test and a second similar one. In the equation given there are n terms such as Σx_1^2, and $\frac{n\,(n-1)}{2}$ terms such as $\Sigma\, 2\, x_1 x_2$, so that we have

$$\sigma^2 \text{of} \left(\frac{x_1 + x_2 + \cdots x_n}{n}\right) = \frac{nN\sigma_1^2 + n\,(n-1)\,N r\,\sigma_1^2}{n\,N} = \sigma_1^2 \left(\frac{1}{n} + r - \frac{1}{n}\, r\right)$$

As n becomes indefinitely large the expression in paranthesis becomes r, so that, representing the standard deviation based upon a very large number of tests, *i. e.* the true standard deviation, by σt we have:

$\sigma_t = \sigma_1 \sqrt{r}$ (Formula giving the standard deviation of the average of an indefinitely large number of similar tests, knowing the standard deviation and coefficient of reliability of the single tests)

In the case of the Kansas Silent Reading tests, taking the reliability coefficient = .25 and the standard deviation σ_1, we have

$$\sigma t = \sigma_1 \sqrt{.25}, \text{ or } \sigma t = .5\ \sigma_1$$

In the original distribution of scores it was necessary to go out from the mean a distance of .215σ to reach a score half way between the median scores of two grades. It will accordingly be necessary in terms of the distribution of true scores to go out $\frac{.215}{.5}$ (=.43) standard

deviations to reach the point midway between grades and to go out .86 σ to reach a neighboring grade median. Therefore according to the overlapping reported by F. J. Kelly there is 33 per cent, instead of 17 per cent, of correct classification and, referring to the frequencies given in a normal distribution, 19.5, instead of 33 per cent, exceed or fall short of the median accomplishment of the grade above or below.

All measures of overlapping thus far reported are incorrect as indicative of the numerical amount of overlapping in true ability The correction in the measure of overlapping here derived is very easy to apply; a table of the frequencies in a normal distribution and the coefficient of reliability being the only new data required.

It is seen that great overlapping is as much an indictment of the test as it is the adequacy of the classification, unless it has been determined that the test itself is highly reliable. Three factors make for overlapping: (a) unreliability of tests; (b) unimportance of function tested; and (c) failure to properly classify, and only in so far as the first two are known not to be the causes can improper classification be established.

TRUMAN L. KELLEY.

Teachers College, Columbia University.

THE USE OF A COMPOSITION SCALE

My experience with the Hillegas Scale has been very different from that of Miss Kate Gordon recorded in the November, 1918, issue of this JOURNAL.

I do not find in my notes complete records of my experiments with the Hillegas Scale, but I do have at hand some results with the Nassau Supplement, and as they are in the main similar they are here given.

A mixed class on Tests and Measurements of 16 students in a summer normal school first rated on the percentage basis a composition, typewritten with all its imperfections. The following day 15 of the same pupils rated the same composition by the Nassau Supplement.

The results of the first rating were arranged into 10 groups so as to be more comparable with the groups of the scale. The results were as follows:

Percentage Ratings %	Frequency	*Nassau Supplement* Score	Frequency
0–10		0	
11–20		1	
21–30		2	
31–40		3	1
41–50	4	4	9
51–60	2	5	5
61–70	8	6	
71–80	1	7	
81–90	1	8	
91–100		9	
	16		15

After a further study of the scale these same students were given a set of 32 seventh grade compositions to rate. Each student had two compositions. After recording the results on the blackboard, the students exchanged sets of papers and then proceeded to rate them again on the basis of the scale. Thus each composition was rated by two students. The results were as follows:

Quality	*First time*	*Second time*
0		
1		
2		
3	5	5
4	13	14
5	8	7
6	4	4
7	2	2
8		
9		
	32	32

In combining the two ratings it was found that—

Both marks the same	16
Difference of 1 step	15
Difference of 2 steps	1

Ernest C. Witham

Southington, Connecticut.

THE CHILD'S USE OF NUMBERS

This paper is an attempt to answer two questions: (1) What arithmetical knowledge and power should be acquired by a child by the end of grade VI? And especially, (2) what should be the content of the arithmetic course in grades IV, V, and VI?

The normal American child is the one considered,—one well brought up in the middle class American home, who reaches grade VII, at about the age of twelve. The needs of such a child will set the standard, whatever be the special provisions required for the abnormal child.

School at best provides an unnatural environment for a child. His out-of-school interests are the best expression of his personality. Accordingly I have gone to the children, who knew me, and asked them to tell me what uses of numbers they make or notice outside of school. The children were members of the Henry L. Pierce School in a choice residence section of the city of Boston. I went to twelve classes of grades IV, V, and VI, and to two immature classes of grade VII; in a most informal and friendly way I got them to give me the information I sought.

THE DATA

I have attempted to tabulate the children's answers. Most of the items reported by a lower grade were also reported by higher grades. I have credited each item to the lowest grade that reported it. Here, then, are the ways in which children of their own accord, use numbers.

A. Games (Regarded by children as of great importance).

Reported first by Grade IV

1. Calling or reading numbers
   (a) Football signals
   (b) Spots on dominoes
   (c) Spots on dice
   (d) Number on sled to indicate size
2. Counting
   (a) Marbles and tops
   (b) To give children time to hide
   (c) Keeping score
   (d) Number of "swings" while taking turns
   (e) Strikes in baseball
   (f) Number of bounces of ball
   (g) Number of children in a game
   (h) While going down hill to estimate length of slide
   (i) Playing with top counting machine

Reported first by Grade VII

   (j) Number of runners in a race

Reported first by Grade IV

3. Playing Store
   All uses of numbers given below under "Going to Store" are also made while *playing* store

Reported first by Grade V

4. Playing Bank
   Matches were used to represent money

B. Going to Store (All use of money is of relatively small importance to child)

Reported first by Grade IV

1. Knowledge of quantity (such as pounds) of articles to be bought
2. Number of articles to be bought
3. Adding, to find total cost of articles bought
4. Counting change
5. Reading number on paper money to determine denomination

Reported first by Grade V

6. Knowledge of price of article to be bought

C. Other Instances of Reading (or Recognition) of some one number

Reported first by Grade IV

1. Telephone number
2. Automobile number
3. Policeman's number
4. Number on fire engine
5. Number on stove
6. Number on street car
7. Number on police station
8. Number of G. A. R. Post
9. Number on house (both in writing addresses and to find a certain house)
10. Telling time
11. Date, year, and day of month
12. Reading thermometer
13. Tire pressure guage

Reported first by Grade V

14. Number on trainmen (elevated trains)
15. Number on firemen
16. Number on newspaper boy
17. Number of regiment
18. Number of submarine
19. Number of pew in church
20. Number of stateroom on boat
21. Number on speed-guage of Victrola
22. Numbers on signal-boards of elevators
23. Numbers in a watch
24. Number by which book is known in library
25. Number of library card
26. Number of record for use on Victrola
27. Numbers in shoes
28. Numbers of floors near doorbell push buttons
29. Number on blinds
30. Number on postage stamps
31. Number on keys and lockers at bathhouses
32. Age of person
33. Numbers in music lessons to indicate what fingers are to be used
34. Date on building
35. Date on monument
36. Numbers gotten by measuring objects
37. Number of any one page of a book

38. Number of page in newspaper to which article is continued from page one
39. Some streets have numbers for names
40. Number as part of name of ruler (Haakon II of Norway)
41. Number of minutes allowed in telephone toll call
42. Page of book at which one stops reading

Reported first by Grade VI

43. Train time in timetable
44. Noting numbers on checks while helping father

Reported first by Grade VII

45. Number on fire alarm box
46. Number on motorman (street car)
47. Numbers on scales (weighing)
48. Number of seat in theater
49. Number of size of rubbers
50. Number on delivery wagon of large store
51. Number on ice cream can
52. Dates in history (Home reading, not for school)
53. Sizes of bits (tools used at home)
54. Total number of pages in book
55. Time when sun rises and sets

D. Other Instances of Counting

Reported first by Grade IV

1. Money belonging to child
2. Trading stamps
3. Postage stamps in child's collection
4. "Checks" good for ice cream at store
5. Books at home belonging to child
6. Stars on flag
7. Stones while digging garden
8. Stars at night

Reported first by Grade V

9. Steps in dancing lessons
10. Days until some holiday
11. Stories in a book
12. Houses on a street
13. Serving of food to supply family at table
14. Time in music lessons and dancing lessons
15. Number of pages read in book
16. Numbering of pages when using typewriter

Reported first by Grade VI

17. Number of Christmas presents
18. Number of guests at table
19. Stripes (as well as stars) on flag

Reported first by Grade VII

20. Money, when selling
21. Stitches while sewing
22. Number of places while setting table

23. Number of times in dumbbell exercises
24. Number of tracks in large railroad station
25. Number of dishes washed
26. Windows, to know how many shades to buy
27. Stripes on policeman's sleeve
28. Newspapers, by boy who is selling them
29. Advertising circulars sent out by father (counted by child)
30. Pieces of carrot for rabbit
31. Rabbits, when shut in
32. Strokes of ax while cutting wood
33. Tulips, and other flowers
34. Number of articles ironed
35. Number of keys on piano (counted for fun, not a part of lesson)
36. Number of strokes one can swim
37. Number of steps required to cross street
38. Number of letters per minute in signalling (boy scout)
39. Number of strokes while beating carpet

E. Miscellaneous

Reported first by Grade IV

1. Adding items in personal account book

Reported first by Grade V

2. Adding up Larkin Soap order
3. Making change while working in store

Reported first by Grade VI

4. Making out bills at store where boy works

Reported first by Grade VII

5. Timing what is cooking

The Results

Thus do the children reveal the uses of numbers that everyday life causes them to make. Compare this mass of data with the table of contents of any textbook in arithmetic and notice how limited the children's uses are. *Reading of numbers and counting include nearly every item recorded above.* Again let us remind ourselves that these are children from good homes, sheltered as we schoolmasters shelter our own children. In this study I am seeking to learn what is true of normal children.

Below grade VII, then, there are no needs felt by the child which require the teaching of arithmetic in the school. All he cares for is to count and to read numbers.

What, then, shall we teach?

In the first place we may at once discard *applied* arithmetic in the grades below the Junior High School. Our textbooks may drop the "Practical Problems," because they are as foreign as Chinese to a child's way of thinking.

But do not think that a child dislikes numbers. I find in these younger children a keen delight in the manipulation of pure numbers (abstract numbers). Mechanical drill is now a pleasure.

Grades IV, V, and VI, then, afford the ideal opportunity for drill in pure numbers (that is, abstract numbers, as opposed to applied arithmetic). This number drill should give the child knowledge and power to work the four processes (addition, substraction, multiplication and division) with integers, common fractions (what little is needed) and decimals. The work should be systematic, but not many reasons should be given for the various processes taught; for example, for inverting the divisor in division of fractions. Logical reasons in arithmetic can be easily taught in the Senior High School.

The work with integers and decimals, by the last part of the sixth grade, should be as difficult as a person will meet until he uses an adding machine and a table of logarithms. The Courtis Tests should come to a maximum in grade VI.

SUMMARY

The children in what we now call the elementary school grades make little use of arithmetic in their life outside of school. They like to count and to read numbers, and that is about all. But they have keen delight in the manipulation of pure numbers. Here, then, we have an opportunity to perfect them in integers and in common and decimal fractions. Then comes the Junior High School when the child awakens to the meaning of life. The arithmetic drill in grades IV, V, and VI, establishes the foundation for the mathematics of the Junior High School.

PHILO G. NOON.

Boston, Massachusetts.

THE ATTENTION VALUE OF LECTURING WITHOUT NOTES

It is a matter of common observation that a lecturer more easily holds his audience when speaking without notes than when reading from them, but so far as is known to the writer no attempt has yet been made to submit this accepted fact to a quantitative statement, or to analyze the factors that may be significantly compared in the two methods of delivery. While the relative advantages of speaking and reading statements to an audience obviously vary widely with individual lecturers, there are unquestionably certain general tendencies which, if successfully analyzed, would be found to obtain with a high degree of constancy for a considerable majority of speakers.

Two types of factors must be clearly distinguished. For one thing the lecturer without notes unconsciously varies his spoken words according to the varying responses from his hearers. He continually modifies set statements into conversational form, changes

his method of emphasis to meet the fluctuations of interest that he sees about him, elaborates where he finds that he has not been clear, and forces himself to greater animation when he meets evidence that he has been dull. These subjective changes in individual speakers are probably too diverse to admit of constant formulation, and would not lend themselves easily to experimental measurement.

But aside from the lecturer's sympathetic variations of statement and gesture there is a sheer difference in the audience according to whether it is listening to reading or to speech. This difference is immediately apparent in the uneasy restlessness of any audience the moment that a speaker who has been expected to talk to them suddenly produces a manuscript. This fact has often impressed the writer as he noted his own responses as a member of an audience, and the following class experiments were made in the effort to arrive at a quantitative statement of the difference in question.

The general method was to give identical material to two very similar groups of students, reading from lecture notes to one class, and speaking from memory to the other. Immediately after hearing the material the students were instructed to write out in the completest detail all they could remember of what they had just heard. Then the papers were scored according to the number of ideas correctly remembered.

The two groups of students numbered 39 and 61 respectively. Both were made up of college undergraduates at the University of Minnesota, and all were beginning students in psychology. Fifty-four per cent of each group were women, in actual numbers 21 of the 39 and 33 of the 61. That the general intelligence of the two groups was approximately the same may be indicated by the distribution of the marks in the course in psychology. The group of 39 numbered 5 A, 7 B, 14 C, 9 D, and 4 E grade students. The group of 61 numbered 6 A, 13 B, 25 C, 11 D, and 6 E students.

Two class experiments were performed, one preliminary and one final. For the preliminary experiment the familiar Marble Statue test was used as material for scoring logical memory. The story was spoken from memory to the larger group and read to the smaller group. The distinctive fact in this preliminary trial was that both groups were aware that the material concerned a special test, and attended with an interest in competition for score. The score for the two groups shows no significant difference. Those who heard the material read averaged 40, with a mean variation of 6.2; those who heard it spoken averaged 41.6, with a mean variation of 3.3. Evidently when the audience is attending competitively the matter of reading or speaking becomes of small importance.

The distinctive feature of the final experiment was that the material for use in the memory test was introduced as part of the regular lecture in psychology, and the students were in no way led to suppose that any peculiar importance attached to it. The material used was a 184 element statement concerning the life and scientific work of Helmholtz, and came at a time when both classes

had just finished the study of visual and auditory sensations in Pillsbury's "Fundamentals of Psychology." The unusual importance of Helmholtz to the subject of psychology was made the justification for introducing a certain amount of biographical material while summarizing the work in audition and vision.

Exactly five minutes were given to the reading and to the speaking, and a long previous practice had assured practically the same kind of emphasis, pauses, and changes of intonation, as heard by the two classes. A person standing outside the room would have had great difficulty in distinguishing which statement was read and which one spoken. But the scores in this test were noticeably different for the two groups. Reading gave an average score of 49.6, with a mean variation of 14.4, and speaking an average of 67.5, with a mean variation of 15.7.

This average difference of 36 percent as contrasted with the negligible difference in the preliminary experiment seems to indicate that the disposition of an audience to give attention is one third greater for spoken utterance than for reading. When any ulterior motive, such as competition, comes into play, attention is about equally effective for reading and speaking, at least for a short time.

It may be concluded that a speaker who reads his address has to assume an unusual earnestness of purpose in his hearers, a much greater earnestness than normally exists in the classroom. And when this sheer difference of 36 percent in the original disposition of the audience is taken in conjunction with the still larger differences due to the reciprocal exchange between speaker and audience, it is hard to escape the conviction that the lecture method, in the sense of the reading of notes to hardly persuaded students, is one of the most dubious features of present day method in college teaching. To the extent that a teacher makes use of notes in the classroom, to that extent he is incurring the danger of reducing his real function in the college to the mere marking of class attendance.

HENRY T. MOORE.

Damouth College.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive researches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3.00 per year of nine monthly issues, payable in advance. Single issues are 50 cents each. Back volumes are $5.00 in original wrappers and $6.25 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X NOVEMBER NUMBER 9

EDITORIAL

THE NATIONAL RESEARCH COUNCIL

The recent World War was preeminently a scientific war. It was a clash of brains as much as a contest of brawn. It was due to the scientific organization and conservation of their resources that the Germans were able to come so near to their goal, and were able to hold out so long after the hope of a smashing success had faded. Trench warfare, the airplane, the long distance gun, the submarine, the tank, are merely the outstanding examples of thousands of scientific devices to sustain the cause and overthrow the enemy. It is distinctly to the credit of the leaders of this country that the significance of scientific organization was realized at the very outset and that every effort was made to mobilize speedily and effectively the scientific resources of the country. This was brought about through the National Academy of Sciences, which at the request of the President established the National Research Council, an organization representing the chief scientific activities of the country. An Executive Order of May 11, 1918, indicates the functions and duties of the Council, in part, as follows:

"In general, to stimulate research in the mathematical, physical and biological sciences, and in the application of these sciences to engineering, agriculture, medicine, and other useful arts, with the object of increasing knowledge, of strengthening the national defense, and of contributing in other ways to the public welfare.

"To survey the larger possibilities of science, to formulate comprehensive projects of research, and to develop effective means of utilizing the scientific and technical resources of the country for dealing with these projects.

"To promote cooperation in research, at home and abroad, in order to secure concentration of effort, minimize duplication, and stimulate progress."

Educational psychology was actively represented on the Council from its very beginning. Already a group of psychologists had rendered signal service in devising tests of intelligence to be applied to the National Army. Early in 1919 a special committee of the National Research Council consisting of Dr. R. M. Yerkes, chairman, Dr. M. E. Haggerty, of the University of Minnesota, Dr. L. M. Terman, of Stanford University, Dr. E. L. Thorndike, of Teachers College, Columbia University, and Dr. G. M. Whipple, of the University of Michigan, secured from the General Education Board a grant of $25,000 for the purpose of working out a series of tests of intelligence for use with school children, similar to those already employed in the army. After detailed study the committee selected some twenty tests and gave these a careful trial on five thousand children. From these tests two series were selected which seemed to give the most satisfaction, and in order to perfect the procedure and to still further check up on the results the tests were tried on several thousand more children. As a result the committee has now before it and will shortly be able to offer to the country two series of group tests of general intelligence of a high degree of reliability that have been carefully tested out on large groups of children, and that give as satisfactory a measure of intelligence as any scale now available. The practical value of such a group scale is tremendous. Schoolmen generally have come to see the advantages of the Binet measurements of intelligence, but the necessity of examining each pupil individually has put such a time cost on the use of the scale as to restrict its employment to special cases. With such a group scale fifty or a hundred children may be examined in an hour, and the only further labor is found in scoring and tabulating the results. We may confidently expect, therefore, that the National Research Council tests will make intelligence testing a routine measure in the schools, and this will be the entering wedge for other studies of pupils' abilities and attainments, and for a radical adjustment of school work to meet the needs of pupils.

J. C. Bell.

NOTES AND NEWS

The New York Society for the Experimental Study of Education held its first meeting of the year in Washington Irving High School, Friday evening, October 10. The general topic of the Meeting was "Problems in the Teaching of English." The following papers were read: "Efficiency of Student Correction of Composition Papers," J. C. Tressler, Newton High School; "The Relative Usefulness of Composition Scales," Sterling A. Leonard, Lincoln School; "Studies in the Appreciation of English Poetry," Professor M. R. Trabue, Teachers College.

The Society for the Promotion of Engineering Education has appointed a committee on intelligence tests for admission to engineering schools. This committee has prepared a plan for cooperative research on admission, and contemplates a series of reports on the predictive value of various tests with respect to subsequent scholarship. A battery of six tests (arithmetic, algebra, geometry, intelligence, physics, and technical information) was prepared for use this fall, and the results of these will be worked over and reported on next summer.

The University of Illinois Bureau of Educational Research has issued an enlarged list of educational tests that it is prepared to furnish, and promises a descriptive list giving detailed information and a brief bibliography of each test. The latter list will be sold at twenty cents per copy.

The National Committee on Mathematical Requirements has received a grant of $16,000 from the General Education Board for the enlargement of the scope of its work. It proposes to make a careful and thorough study of the teaching of mathematics, and to report on various reforms that have been advocated. Professor J. W. Young, of Dartmouth College, and Mr. J. A. Foberg, of the Crane Technical High School, Chicago, have secured leaves of absence from their institutions and are devoting their entire time to the work.

Tests of intelligence are being increasingly used to determine fitness for school work. The school authorities of Lawrence, Kansas, have decided to admit to the schools those children only five and one-half years old who show a mental age of six years, and at the University of Chicago kindergarten children who stand high in intelligence are found to do well in the work of the first grade. These are indications of the growing disposition to study the child and to fit the school activities to his needs and capacities.

In an address before the London Modern Language Association Mr. E. Allison Peers advocated the formation of a department of educational experiment and research whose function it should be to carry on cooperative investigations and publish data on all kinds of innovations in method. After the address a large number of teachers expressed their desire to take part in the experiment.

The Alabama legislature has passed the Alabama Mental Deficiency bill appropriating $200,000 for the establishment of the Alabama Home for Mental Inferiors at Tuscaloosa.

Warwick and York announces the publication of an Introductory Psychology for Teachers by Professor Edward K. Strong, of the Carnegie Institute of Technology. This text has been developed along novel lines, and in the opinion of many who have examined the work will better meet the needs of prospective teachers than any book now on the market.

At Harvard University William McDougall, of Oxford University, has been elected professor of psychology to fill the vacancy caused by the death of Hugo Munsterberg. Professor McDougall will take up his work at the beginning of the next academic year. Dr. Herbert S. Langfeld has been appointed director of the psychological laboratory, and Drs. L. T. Troland and Floyd H. Allport have been appointed instructors in psychology.

Professor Daniel Starch, of the University of Wisconsin, is on leave of absence for the first half of the present year. He is giving a three-hour course of lectures during the semester at Harvard University.

Mr. Leon O. Smith, director of research in the Omaha Public Schools, has been advanced to the position of assistant superintendent.

Dr. Walter S. Monroe, of Indiana University, has been made assistant director of the Bureau of Educational Research of the University of Illinois..

J. M. O'Gorman, lecturer in education at the University of Illinois, has accepted the position of professor of education and psychology in the Montana State College.

PUBLICATIONS RECEIVED

ERNEST R. BRESLICH. *Correlated Mathematics for Junior Colleges.* Chicago: The University of Chicago Press, 1919. Pp. xiii, 301. $1.25.

This constitutes the author's fourth-year text in correlated mathematics. Instead of the traditional algebra in the first year, plane geometry in the second, solid geometry and trigonometry in the third, and analytical geometry in the fourth, this series of texts presents the simpler aspects of the whole subject of mathematics in the first year, and proceeds with increasing complexity through the subsequent three years. The series is constructed throughout from the standpoint of the psychology of the pupil, and gives a well-rounded conception of elementary mathematics.

B. R. BUCKINGHAM. *Extension of the Ayres Spelling Scale.* Bureau of Educational Research, University of Illinois, Urbana, Illinois. 1919. Ten cents.

This extension of the Ayres Scale will distinctly increase its value and usefulness. The 1000 original Ayres words are retained and printed in Roman type. To these 505 new words have been added in italics. The new words were selected from spellers, and their difficulty determined by tests in several states. Most of them fall at the upper end of the Ayres Scale, and it was found desirable to extend the scale six columns to the right. In addition scores for the ninth grade are included for all columns to the right of P. This is the kind of work that will be of real service to educational procedure. The Scale should be still further extended to include perhaps five thousand words.

STUART A. COURTIS. *The Gary Public Schools. Measurement of Classroom Products.* New York: General Education Board, 1919. Pp. xxii, 532. Thirty cents.

This is the final volume of the report on the survey of the Gary schools. It is not only a detailed account of what is unquestionably the most accurate experimental survey of a school system ever made, but is also a critical and constructive discussion of the whole subject of educational measurements. The specific studies reported are handwriting (scored by the Ayres scale), spelling (compared with the Ayres standards), arithmetic (measured by the Courtis tests, Series B, and the Cleveland Survey tests), composition (evaluated by the Hillegas scale), and reading (measured by the Gray Oral Reading scale and the Kansas Silent Reading tests). No student of educational measurements can afford to be without this volume.

KARY C. DAVIS. *Horticulture.* Philadelphia: J. B. Lippincott Company, 1919. Pp. vii, 416. $1.75.

Here is a splendid text for high and normal schools, which combines instruction in gardening, orcharding and small fruit culture, covering in a comprehensive yet wholly practical manner the entire subject of horticulture. The abundant illustrations are carefully planned to supplement and clarity the discussions of the text, and the questions at the end of each chapter stimulate the student to reflective thinking. The book on account of its simple and easy style will also make a strong appeal to the practical home gardener.

Elbert Kirtley Fretwell. *A Study in Educational Prognosis.* New York: Teachers College, Columbia University Contributions to Education, No. 99, 1919. Pp. 55.

The chief problem in this study was what tests are of most value for educational prognosis? The work was carried on under the supervision of Professor Thomas H. Briggs in the Speyer School, New York City. Eleven tests were given to seventy-four 6 B boys at the beginning and again at the end of the school year, and the results of these were compared with the class records and the judgments of teachers as to school abilities. The author concludes that the tests afford a better means of educational prognosis than do previous school marks or teachers' estimates. The order of value of the tests is reading, visual vocabulary, opposites, spelling, completion, arithmetic, easy directions, mixed relations and composition.

Walter S. Hunter. *General Psychology.* Chicago: University of Chicago Press, 1919. Pp. xiii, 351. $2.00.

By the term General Psychology the author intends to indicate his conviction that the field of psychology is much broader than normal adult psychology, and that an introductory work should serve to give the student a bird's eye view of the entire subject. One-third of the book is, therefore, devoted to the consideration of the fields of psychology, including animal psychology, individual and applied psychology, abnormal psychology, and social and racial psychology. In the remainder of the book, the section on normal adult psychology, the effort is made to bring out the essential results of experimental psychology by means of a somewhat detailed account of representative experimental studies. There are many excellent figures of the nervous system and the sense organs, and several graphs showing the results of experiments. As a concise and well balanced introductory text the book deserves careful consideration.

Buford Johnson. *Practice Effects in a Target Test—A Comparative Study of Groups Varying in Intelligence.* Reprinted from the Psychological Review, Vol. 26, 1919. 300-326.

At the New York State Reformatory for Women three groups of inmates were formed on the basis of intelligence as shown by the Stanford-Binet tests and the Yerkes-Bridges scale. The average I. Q. for each group was 99, 73, and 56. Twenty-seven daily practice periods of ten trials with each hand were given to each group. The high intelligence group had the highest initial and final average, the medium group was second, and the low group last. The medium group, however, showed more improvement than either the high or the low group. The individual learning curves suggest temperamental differences as more significant than differences in intelligence per se.

Wilfrid Lay. *The Child's Unconscious Mind.* New York: Dodd, Mead and Company, 1919. Pp. 329. $2.00.

In this book the author tries to show the relations of psychoanalysis to education. The entire argument is based on the assumed activities of the unconscious (sub-conscious and co-conscious are used as synonyms) mind. This is nowhere clearly explained but under the name of unconscious wish, unconscious inference, unconscious memory, etc., is made responsible for most of our attitudes and strivings. Thus the author goes on about the interplay between the conscious and the unconscious,

identifications, projections, introjections, compensations, sublimations, and the whole stock in trade of the psychoanalysts, drawing his illustrations from child behavior and applying his conclusions to educational practice. All this, shot through as it is with the usual crimson thread of sex reference, impresses the critical psychologist as an erotic phantasmagoria, punctuated here and there with shrewd observations on child life and pertinent suggestions for school room procedure.

Nevada, Report of a Survey of the University of. Washington: Bureau of Education, 1917, No. 19. Pp. 184. Twenty-five cents.

The report refers to the suspicion and distrust that are rife among the people of Nevada regarding the policies of the University, examines its present organization and equipment, commends much that is found, and makes suggestions for betterment. Among the recommendations the increase of attention to the training of teachers for Nevada schools is given a prominent place.

HELEN RICH NORTON. *Department Store Education.* Washington: Bureau of Education, 1917, No. 9. Pp. 79. $.15.

An account of the course of study, methods of teaching and follow-up work developed at the Boston School of Salesmanship under the direction of Lucinda Wyman Prince. There is also a chapter on the teaching of salesmanship in the public schools.

STEPHEN PAGET. *Adolescence.* New York: E. P. Dutton and Company, 1919. Pp. 46.

The address of an eminent physician to Oxford University Extension Students in 1917. A splendid instance of the way educated English people are apt to talk around this subject without ever coming to grips with it.

Personnel Work in the United States Army. Washington: Adjutant General's Office, 1919. Pp. 15.

This pamphlet describes the classification of soldiers according to their industrial qualifications, lists the trade tests developed for determining industrial skill, and presents the rating scale for the classification and rating of officers.

FRANCIS G. PEABODY. *Education for Life.* New York: Doubleday, Page and Company, 1918. Pp. xxiv, 393. $2.50.

The story of Hampton Institute, told in connection with the fiftieth anniversary of the school. It tells of the negro in the Civil War, in the reconstruction period, and in the period of industrial expansion in the South. It recounts the heroic efforts of General Armstrong to provide educational opportunities for the negro, and to raise money for the support of Hampton Institute. The book is profusely illustrated and is provided with valuable statistical appendices. Written as it was by the vice-president of the board of trustees, it may be considered the official history of the first institution for the higher education of the negroes.

JOSEPH PETERSON AND QUENTIN J. DAVID. *The Psychology of Handling Men in the Army.* Minneapolis: Perine Book Company, 1918. Pp. v, 146.

This book is an attempt by a professor of psychology and a lieutenant in the army to apply the psychology of group activities to the problems of handling soldiers. It consists largely of anecdotes drawn from military and athletic life illustrating the values of competition, play, teamplay, leadership, habit and discipline, loyalty, and the principles of learning in dealing with recruits.

VOLUME X DECEMBER, 1919 NUMBER 10

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

ADULT TESTS OF THE STANFORD REVISION APPLIED TO COLLEGE STUDENTS

Helen Hubbert Caldwell, Ph. D.

University of Wisconsin

SINCE the launching of investigations regarding intelligence tests and college students by Cattell (1), in 1896, a number of such studies have been made. Many of these were on freshmen only, and not all of them bear directly upon the present undertaking. Only those in which a comparison is instituted between results obtained from the tests and college standing as shown by grades, or estimated intelligence, will be mentioned here.

Wissler (2), discussing Cattell's results reports the following correlations of performance in the tests with University grades:—

Reaction time	-.02
Marking A's—	-.09
Association—	.08
Color naming—	.02
Logical memory—	.19
Auditory memory (position)	.16

It will be seen that the average correlation of the tests used with college grades was .09, which is so low as to be negligible and apparently justifies the conclusion that, "while the marks of students" (in different subjects) "correlate with each other to a considerable degree, they show little tendency to do so with the mental tests of the psychologist."

Calfee (3), using "Four General Intelligence Tests" on one hundred and three freshmen at the University of Texas, found these correlations with grades:—

Card sorting—	.32
Card dealing—	.25
Alphabet dealing—	.17
Mirror test—	16

The average here is seen to be approximately .23.

Rowland and Lowden (4) in their attempt to determine the relative value of different groupings of tests given to students of Reed College, found that the highest correlation between any group of tests and college grades was .37, +.06 (Yule's formula). The tests selected, after discarding those which individually showed low correlations, were opposites, logical memory, rote memory, judgment and attention.

Waugh (5) gave seven tests to students entering Beloit College and correlated their performances with class standing. Correlations for the calcellation of A's, designed to measure concentration of attention, and for the puzzle box, which measured ingenuity, are not given since they were below .2. Correlations of the remaining tests with class standing appear below. Those relating to motor control are purposely omitted.

Association (opposites)	.54
Speed of learning (substitution)	.24
Range of information	.47
Memory (connected passage)	.40

The average is seen to be .41. Waugh's correlations for association and memory are much higher than Wissler's; his results from the attention test were evidently below .2 whereas it is one of the tests chosen by Rowland and Lowden for its high correlation. It would be difficult to determine whether such variations are due to inherent differences in the tests, or to differences in the methods of applying them, since neither tests or methods conform to any general standard.

Bell (6) used nine tests on 750 freshmen. His method differed from those preceding in that he adapted the tests for use with *groups*, the average time required to take all nine tests being forty-five minutes. The tests given were standard and well known.

Cancellation of triangles (after Simpson)
Addition (Simpson)
Association (adaption of Learning Pairs of Simpson and Thorndike)
Recognition
Selective judgment (Bonser)
Directions No. 1 (Woodworth and Wells)
Directions No. 2 (Woodworth and Wells)
Alternatives (Squire)
Completion (Terman and Childs)

The highest correlation obtained between class marks and any test score was .31 for the marks in English and the Completion test.

Kitson, (7), working with forty students in the College of Commerce and Administration of Chicago University, used many more tests than previous experimenters, sixteen in all, and found the correlation between standings in University grades for the year and standings in the tests to be .44. His tests were:—

1. Number checking (Woodworth and Wells)
2. Numbers heard (Whipple)
3. Objects seen
4. Logically related material, Heard
6. Logically related material, Seen

5, 7, 8, 9. Logically related material Heard and Seen, deferred reproduction.

10. Opposites (Woodworth and Wells)
11. Constant Increment (Woodworth and Wells)
12. Hard Directions (Woodworth and Wells)
13. Directions—Oral
14. Word-building (Whipple)
15. Sentence-building (Whipple)
16. Business Ingenuity.

These same tests used a year later with Freshmen gave a correlation with grades of only .20, which leads Kitson to conclude that "these results give point to the remark that the correlation between intelligence as measured by tests, and academic standing is not so high as is popularly supposed." A correlation of the test results with estimated intelligence, intelligence being defined as *ability to adapt one's self to new situations,* and estimated by the dean on the basis of the grades, but with liberty to change the ranking as he thought best, gave a result of .57 (P. E. .05). Kitson seemed not at all surprised at either correlation, since "an analysis of the conditions of University life shows that many other factors besides intelligence enter in to determine class-room standing."

King (8) gave the Completion, Hard Opposites, Logical Memory and Lanes Tests to students at the University of Iowa, obtaining a correlation of .41 between the results of the tests and class standing as estimated by certain instructors. He concludes that "in these tests we have measures of ability which should be of considerable value to advisory officers in a University."

King and M'Crory (9) used the following tests on about five hundred Freshmen at the University of Iowa:

1. Courtis Standard Arithmetic Test-Series B
2. Mixed Relations (Whipple set)
3. Opposites (Simpson, rearranged by Gold and King)
4. Completion Test (Simpson)
5. Visualization (three inch cube test with each of the nine questions weighted and assigned a percentage value).
6. Range of Information (Whipple)
7. Logical Memory

Correlations are reported separately for boys and girls, and inter-correlations of tests are given as well as the correlation of each test with college grades. Combining the records of boys and girls, the average correlation of grades with the tests used seems to be .39. Of course this is only approximate, since to get an accurate coefficient, the boys and girls would have to be re-ranked together.

Dr. Starch told the writer that in correlating the results of the Stanford Adult tests with University standing (University of Wisconsin), and with estimated intelligence, the following coefficients of correlation were obtained:—

Test results with grades	.30
Test results with estimated intelligence	.75

Report of this investigation has not yet been published.

Miss Downey's work (10) deals with the Adult Intelligence Tests standardized by Terman, which were the ones used by the writer, and is therefore of more specific interest in this connection. The XVI and XVIII year old tests were given to thirty faculty members, thirty-two upper classmen and thirty-two freshmen, with these results; only one member of the faculty fell below the superior adult level as measured by the intelligence quotient, fifty percent fell within the very limited range of 110.4 to 113; two thirds of the upper classmen were in the superior adult group, one third in the average adult group, and fifty percent had I Q's between 104.1 and 113; of the freshmen, one third (approximately) were of superior, one half of average and one fifth of inferior intelligence. It is significant that all but one of the inferior group dropped out of college after the first year. Fifty percent of the freshmen ranged from 101.5 to 107 in I Q's.

The most surprising thing about Miss Downey's results is the massing of intelligence quotients in so limited a range for each group tested. Such a massing suggests the conclusion that the groups chosen for the tests were unusually homogenous, and seriously invalidates any conclusions from correlations, since it would be

impossible to eliminate the chance element in ranking say 15 persons between 110.4 and 113. Miss Downey herself partially recognizes this fact. "In conclusion, I would emphasize the value of the Stanford Adult Tests as a direct aid in analysis of an individual's make-up. Within a closely selected group, this value would outweigh that of expression of the results in terms of an Intelligence Quotient."

No attempt was made to correlate the test results with University grades, but a correlation of Miss Downey's ranking of her psychology class with the results of the tests gave .527 (P. E. .07). She believes that the agreement would have been closer but for error in her ranking due to an over-emphasis of the *verbal* factor, and goes on to say, "undoubtedly the most important source of error in judging intelligence is the undue emphasis laid upon verbal fluency as a social measure of general intelligence." We are at once reminded of Goddard's (11) warning that the higher grades of mental deficiency often pass for normality because of the verbal fluency of the persons affected, and it seems likely that the same error may be made in judging normal persons as of average, superior or very superior intelligence. Indeed we are left no room for doubt when we are told that faculty rankings (estimated intelligence) which failed to show any correlation with test results, gave a coefficient of correlation of .55 (P. E. .10) when compared with the results of the vocabulary test alone.

Finally, the tests were divided into three groups, *verbal, digit-span and constructive.* Those reagents of a literary turn of mind excelled in the verbal tests, those scientifically inclined scored higher on the constructive group, while results of the digit-span series showed no difference between groups. If further investigation with larger groups confirms these results we have some basis upon which advisory officers may build in helping the student select his courses.

*TABLE I a

Correlation of Results of Tests with College Grades

Wissler—	.09
Calfee—	.23
Rowland & Lowden—	.37
Waugh—	.41
Bell—	.31
Kitson—	.44
King & M'Crory—	.39
Caldwell—	.44

TABLE I b

Correlation of test results with estimated intelligence

Kitson—	.57
King—	.41
Downey—	.527
Caldwell	.47

The present investigation was made in 1916-17 while the writer was a teacher of psychology at Randolph-Macon Woman's College. Six members of the class on Mental Tests, all seniors, were trained in the technique of the Stanford Revision. Each girl tested a number of subjects under the direct supervision of the author, and had careful instructions in scoring. Not until all were thoroughly familiar with the use of the tests was the work here reported begun. Forty-eight students, twenty-four Sophomores and twenty-four Juniors were used as subjects. The XVI and XVIII year old Terman Tests were used, and all alternates were given, regardless of the performance in the original tests.

For convenience, the Average Adult Tests (XVI years) will be referred to as I-1, I-2, I-3, I-4, I-5, I-6, Al 1 and Al 2, and the Superior Adult Tests (XVIII years), as II-1, II-2, II-3, II-4, II-5, II-6. Careful record was kept of the nature of each wrong response given by each subject. So far as was possible, a verbatim record of the answers was taken. This made it possible to rank, with a fair degree of accuracy, students whose intelligence quotients were the same. I. Q's were computed after Terman (12) by dividing the age, as determined by the tests, by 16, which is taken as the adult age. Each score sheet was inspected by the writer immediately after the completion of the test so that any obscure points might be cleared up while the matter was fresh in the mind of the examiner who had acted as experimenter. It may be added that no experimenter was given a subject with whom she was on familiar terms, lest such familiarity invalidate the results.

Next, four members of the faculty and two students ranked the subjects tested as to intelligence, the latter being defined by general consent as, *general ability to adapt oneself to practical situations.* The average rating for each girl was taken as the basis for compu-

*It may be of interest to know that use of the Yerkes-Bridges Point Scale by Garrison, at the Peabody College for Teachers yielded a correlation of test results with grades of only .21. There is no statement that the Yerkes-Rossy scale for adults was the one used. Yerkes and Burt report relations between results of tests by the Yerkes-Rossy scale and a set of psychology grades, using the method of tertiles, claiming a positive correlation, but giving no coefficient.

tation. Lastly, the average grade of each girl since entering college was calculated from the registrar's record books, grades in all courses for each year being made the basis of calculation. Correlations were then obtained by the Pearson Rank Method, for the two groups (Juniors and Sophomores) separately, and for all subjects ranked together. These correlations are given in the form of a table below:—

TABLE II

Correlations of Results of Stanford Adult Tests with College Grades.

Juniors	
Intelligence quotient and grades—	.69
Intelligence quotient and estimated intelligence—	.46
Estimated intelligence and grades—	.52
Sophomores	
*Intelligence quotient and grades—	.17
Intelligence quotient and estimated intelligence—	.50
Estimated intelligence and grades—	.38
Juniors and Sophomores—Re-ranked in One Group	
Intelligence quotient and grades—	.44
Intelligence quotient and estimated intelligence—	.47
Estimated intelligence and grades—	.37
*Omission of three widely varying records here gives	.48

It is a recognized fact that no intelligence tests are complete without some knowledge of the character and environment of the subjects tested. Such knowledge often gives the key to discrepencies which appear when the data are handled mathematically. There is grave danger that in the attempt to standardize tests and evaluate results the more subtle factors entering into success or failure in performance will be overlooked.

In the Sophomore Group, the low correlation of Intelligence Quotient with Grades is directly traceable to the records of three students. Number one was ranked one and a half in the tests and fifteenth in grades. She was a girl whom everybody liked, and was naturally bright, but her popularity together with a natural disinclination toward exertion caused her to do little studying. But for her native ability she could not have maintained herself in college. Number two, also bright and peculiar, was engaged in

*The author is not prepared to enter into a discussion regarding the relative merits of the Stanford Revision and the Point-Scale, but either of these is preferable to a set which is *as a whole* unstandardized, even though norms for the separate tests have been carefully worked out and standardized.

almost every outside activity which the college afforded, greatly to the detriment of her class standing. She was allowed to return for her Junior year on probation because of her popularity and promise. Number three was an excellent talker, and never failed to recite whether she knew much about the subject or not. This is perhaps the best example in these results of over-rating due to verbal fluency mentioned by Miss Downey. The girl ranked fourth in grades, given largely because of her readiness to recite and her general "well-meaningness" as one student described it, but was twenty-fourth in the tests. When these three records are omitted, the correlation between intelligence quotient and grades for the Sophomores is .48.

This correlation while somewhat closer than any reported above, still lacks much of being as marked as that of the Juniors. There may be several reasons for this difference. The Juniors are, as a rule, steadier, less erratic, less given to outside distractions than the Sophomores, and put forth more uniform effort in daily work. Hence their grades more exactly represent their ability, and correlate more closely with the net results. Again, the grades from which the Sophomore rankings are obtained are almost entirely those of required subjects, while in the Junior rankings are included a full year of electives. As a rule, students are more interested and hence do better work in subjects which they elect than in subjects which are required, and here again the Junior grades would be a more accurate index of ability than the Sophomore grades. It is also quite probable that the somewhat greater maturity of the Juniors plays its part, the age range here being eighteen to twenty-three, while in the Sophomore group it is seventeen to twenty.

In the re-ranking of the two groups as one, even though the three records discussed above are included, the coefficient of correlation between the grades and intelligence quotient is higher than any yet obtained except that of Kitson. When the complexities of the grading system in the average college are considered a closer correlation is scarcely to be expected.

The same sources of error apply in estimating intelligence as in giving grades, and in every case where there is wide variation in rank between estimated intelligence and test results, the reason is to be found in a close study of the character of the subject, as, for example, in the case of the student who was commonly very quiet, and gave the impression of being quite ordinary because "she

never had anything to say for herself," but who had the gift of rising to emergencies such as written tests and examinations, and who ranked second in grades, sixth in the tests, but twentieth in estimated intelligence.

Comparing the coefficient of correlation between the test ratings and estimated intelligence with those obtained by previous experimenters shows it to be higher than King's, but lower than Kitson's or Miss Downey's. The two last, however, reported estimated intelligence on the basis of one person's judgment only, while the ranking in King's report was done by two or three instructors and in the present research by six persons.

It was at first a matter of some surprise to the author that there was no evidence of closer correlation between estimated intelligence and grades, since grades must certainly be taken largely as a basis of estimate, certainly on the part of an instructor. But when the definition of "intelligence" used in making the ranking is considered, one sees that there may be many things besides classroom work which strongly influence the estimation of intelligence, and there may have been also a conscious effort on the part of the faculty members not to allow the high or low standing of the students in their particular classes to exert too strong an influence on the ranking—an effort which may easily have resulted in too low or too high a ranking, as the case might be. Unfortunately there are no reported correlations with which this coefficient may be compared.

Turning next to the results of the separate tests, we find in Table III the percentage of failures on each test for each class.

TABLE III

Sophomores		Juniors	
Test	% failures	Test	% failures
I-1	33⅓	I-1	12½
I-2	12½	I-2	12½
I-3	25	I-3	12½
I-4	37½	I-4	37½
I-5	33⅓	I-5	29 1-6
I-6	29 1-6	I-6	29 1-6
A1 1	58½	A1 1	62½
A12	83⅓	A12	62½
II-1	70	II-1	50
II-2	25	II-2	42
II-3	42	II-3	12½
II-4	46	II-4	33⅓
II-5	54	II-5	70
II-6	50	II-6	62

The fact that in both the sophomore and junior years the alternate tests for the Average Adult Group show a much higher percentage of failures than the regular tests, would seem to throw some doubt upon either their value or their proper placing, and to suggest that if used at all, they perhaps belong in the Superior Adult Group.

Terman himself criticises these tests, Alternate *one* as "too mechanical to tax very heavily the higher thought processes," while of Alternate two he says, "These problems - - - yield interesting results when properly given, but are not without their faults. Sometimes a V S (very superior) subject fails, while occasionally an I (inferior) subject unexpectedly succeeds."

This was exactly the experience of the writer with these tests, but she did not find 60 to 75 percent of average adults passing, and few if any superior adults failing, as did Terman, for 50% of the failures on AI 1 and 46% of those on AI 2 were by students whose I Q was above 110.

The results of this study raise one more question. It is as to the value of the tests classed by Miss Downey as *digit-span tests*, which we shall designate *pure auditory memory tests*. A number of students who ranked high in both estimated intelligence and grades, missed only those questions in the tests involving auditory memory, i, e, I-5, AI 1, II-3 and II-5, thus falling far below in their test rankings. Whatever may be the value of the test for immediate memory, is it not overdoing it somewhat to have three such tests in the determination of adult intelligence, and two of them in one group of six tests, thus constituting 1-3 of the score for that group? Granting that the process of giving the digits in reverse order may possibly involve more than mere immediate memory, that must still be involved, and it seems to the writer that it would be better to replace at least one of these tests with a requirement involving some other ability.

Attention is also called to the fact that the highest I Q possible of attainment by an adult is 121.8. This is near the lower limit of Very Superior Intelligence as designated by Terman, and makes the determination of an adult genius or near-genius impossible. Something should be done to make the scale flexible enough and extensive enough to apply to upper as well as lower limits of intelligence. This is a fault, if fault it be, common to all intelligence tests, that they select inferior individuals much more readily than they do superior ones. From a vocational standpoint this is a less

costly fault than the reverse would be, but it is highly desirable from both the educational and the vocational point of view that there be some better way of selecting those persons who will prove most worthy of intensive and specialized training.

Conclusions

1. The correlation of the test results with scholastic standing is such that the Adult Tests may be useful to advisory officers in diagnosing cases of unsatisfactory scholarship, whether among Freshmen or Upper-classmen. Indeed the value of the tests in this particular would probably increase with advance in standing. A further use of the tests might be made in colleges where a long waiting list is the rule, in order to weed out undesirable candidates for admission.

2. Since the correlation of these test results with grades is as high as that found with any other set of tests, (see table Ia), it is to be recommended that the *Stanford Adult Tests, which have the advantage of being standardized, be used in investigations regarding adult intelligence in order that there may be a common basis for the comparison of results.

3. The tests need revision in a few particulars. The alternate tests for the Average Adult are either wrongly placed or should be changed, and there seem to be relatively too many pure auditory memory tests in the Superior Adult Group.

4. The Intelligence Quotients for adults have less meaning than those for children, and there should be some means for denoting higher adult I Q'S.

5. I Q's must always be interpreted in the light of character study, and standardized character tests in addition to these adult tests are highly desirable.

6. A productive field for further work along this line would lie in the application of the Stanford Adult Tests to persons selected for their brilliancy, in order to determine—
   a. the possible limits of performance
   b. the percentage of failures in each part of the test
   c. The extent to which Miss Downey's division into Literary and Scientific groups on the basis of performance may be expected to hold.

BIBLIOGRAPHY

1. Cattell, J. McKeen, and Farrand, Livingston. *Physical and Mental Measurements of the Students of Columbia University.* Psychol. Rev. 3, 1896, 618-648.
2. Wissler, Clark. *The Correlation of Mental and Physical Tests.* Psychol. Rev. Mon. Sup. 3, 1901, No. 6, 1-61.
3. Calfee, M. *College Freshmen and Four General Intelligence Tests.* Jour. of Educ. Psychol., 1913, 4, 223-231.
4. Rowland, E. and Lowden, G. *Report of Psychological Tests at Reed College.* Jour. of Exp. Psychol, 1916, 1, 211-217.
5. Waugh, Karl T. *A New Mental Diagnosis of the College Student.* New York Times, Mag. Sup., Jan. 2, 1916.
6. Bell, J. Carleton. *Mental Tests and College Freshmen.* Jour. of Educ. Psychol., 1916, 7, 381-399.
7. Kitson, H. D. *Scientific Study of the College Student.* Psychol. Mon., 1917, 23, No. 1, p 81.
8. King, I. *The Relationship of Abilities in Certain Mental Tests to Ability as Estimated by Teachers.* School and Soc., 1917, 5, 204-209.
9. King, I. and M'Crory, Jas. *Freshman Tests at the State University of Iowa.* Jour. of Educ. Psychol., 1918, 9, 32-46.
10. Downey, June E. *The Standard Adult Intelligence Tests.* Jour. of Delinq., 1917, 2, 144-155.
11. Goddard, H. H. *Feeble-mindedness; its Causes and Consequences.* Mac-Millan Company, 1914.
12. Terman, Lewis M. *The Measurement of Intelligence.* Houghton-Mifflin Company, 1916.

A TEST IN LATIN[1]

Daniel Starch

University of Wisconsin

Preparation of the test. There are at least four desirable conditions which an educational test or scale ought to satisfy as nearly as possible. (1) The test or scale for a given subject should measure the knowledge or skill acquired wherever proper instruction or training in that field is provided. The material of the test should be based upon the material that has actually been covered or should be covered in the subject. (2) The test ought to permit of accurate, impersonal evaluation. (3) It ought to permit of duplication or repeated application almost indefinitely. (4) It ought to measure adequately the knowledge, skill or ability for which it is designed. The present Latin test has been prepared with reference to these four criteria.

The first criterion may be satisfied rather easily in preparing a Latin test since the content of Latin in the high school is more nearly standard or uniform than it is in most other subjects. Excepting certain variations in the first year of Latin, the classical texts studied are the same wherever Latin is taught, namely, Caesar, Cicero, and Virgil.

The test as finally constructed is composed of two parts, a vocabulary test and a translation test. The vocabulary test is composed of 100 words selected from Lodge's "The Vocabulary of High School Latin" by choosing every 20th word. Lodge's Vocabulary list contains the 2000 words occurring in Caesar, Cicero, and Virgil.

The translation test is composed of four parts, one for each of the four years of high school Latin. The first year test is composed of twenty sentences selected from five widely used first year tests. Four sentences were chosen at uniform intervals from each of the five texts.[1] For example, Pearson's Essentials of Latin contains 857 sentences. If the 98th sentence was the first one chosen, the other three would be the 312th, the 526th and the 740th. That is,

---

[1] The preparation of the test itself and the compilation of the results were carried out by Mr. J. M. Watters under the direction of the writer.

[2] Pearson's Essentials of Latin; Smith and Lange's First Latin Lessons; Collra and Daniel's First Year Latin; Bennett's First Year Latin; and D'Ooge's Latin for Beginners.

they were chosen at intervals of 214 sentences. The only restrictions observed were that the sentences must contain no less than three words at least two of which must be of different parts of speech. In like manner four sentences were chosen from each of the other four texts. The only difference was that in the texts containing varying numbers of sentences, the interval varied accordingly. For example, D'Ooge's Latin for Beginners contains 490 sentences. The four sentences were selected at intervals of 122 sentences. The sentences from the different texts were then arranged alternately.

The seven sentences for the second year test were selected from Caesar at intervals of twenty-two chapters with the exception that Chapters I and XIV of Book I and Chapter XVII of Book IV were omitted because the first one is frequently memorized and the others present unusual difficulties of translation. The only further restrictions observed in the selection of the sentences were that they must contain no less than six words nor more than thirty-five. In case a sentence failed to meet these restrictions the sentence preceding or following was chosen. For the particular test here used the second sentence of Chapter IV of Book I was the first one selected. The other six sentences were obtained by selecting the second sentence of every 22nd chapter thereafter as just mentioned.

The seven sentences for the third year test were chosen at regular intervals from the four orations against Cataline, the oration for the Manilian Law and the oration in behalf of Archias. The four orations against Cataline and the one in behalf of Archais are approximately equal in length while the one for the Manilian Law is about twice as long as any of the others. This oration was divided into two parts of twelve chapters each. Subject to the same reservations followed in connection with the Caesar test, the seven sentences were chosen by taking the fourth sentence of the third chapter of each of the four orations against Cataline, the oration for the Manilian Law, and each of the two parts of the oration in behalf of Archias.

The ten sentences for the fourth year test were chosen from Virgil at intervals of 500 lines beginning with the 88th line. The first sentence could not be selected from among the first fifty lines since these are frequently memorized.

The idea governing the selection of this particular amount of material for each of the four translation tests was to limit it so

that it could be done as a test within a class period. The aim was to limit the amount for each year test to approximately 125 to 150 words.

Comparison of Test No. I and Test No. II in First Year Latin Percentage Scores

	No. of Words	First 100 Sentences
Test I	122	84.3
Test II	139	83.0

Comparison of Tests No. I and No. II in Caesar

	No. of Words	Percentage Scores
Test I	154	72.9
Test II	142	76.2

The second desirable feature of an educational test is accuracy and objectivity in the evaluation of the result. The specifications for scoring the vocabulary test are as follows:

> Determine the number of words correct, count each word as either right or wrong. Any accepted meaning is considered correct. However, a meaning such as *four* for *quartus* is considered incorrect.

Two methods of scoring the vocabulary test were tried out and compared. One was the plan here adopted, namely, of having the pupils write the English meaning of the Latin word. The other was the matching plan which consisted of presenting the Latin words on one side of the page and the English equivalents numbered and arranged in alphabetical order on the other side. The task of the pupil was to write the number of the English equivalent after each Latin word. The two methods were employed with a class of fifteen students with the outcome that the former plan was adopted. The two methods produced substantially the same scores. The average score by the defining method was 26.4 and by the matching method 23.7. The second or matching plan has the advantage of being entirely impersonal and objective, but it has the disadvantage of requiring too long a time on the part of the pupil to do the test and also of using meanings which for particular Latin words may be unfamiliar to the pupil. The defining method involves to some extent the element of judgment as to whether the meaning given by the pupil is satisfactory. However, this element is slight from the practical viewpoint. The defining method is on the whole the more desirable one.

The translation test is scored as follows:

> Consider the translation correct if the thought is rendered correctly. A word may be rendered incorrectly in two ways, either in meaning or in its relation to other words, each counting as a half point. For example: "His

rebus cognitis" may be six half points, three in meaning and three in grammar. "These things having been learned," or "After he had learned these things" would be correct translations and would receive the full credit of three points or words. But in the translation, "You learned the things from these" each word is grammatically wrong even though the meanings of the words individually are correct. Hence, it will receive three half points or one and one-half points. After determining the total number of points made in the test for a given year, compute the percentage of these points on the total number of the words in the test. For example, a pupil who made 83 points on the first test out of a total of 122 possible points or words would receive a score of 68 per cent.

This method of scoring seemed the most adequate in that it takes account of the meanings of words as well as of their grammatical relations and as such has not only the advantage of accurately evaluating the work of the pupil but also of locating precisely the weakness of the pupil, that is, whether the pupil is deficient in the knowledge of words or in the knowledge of grammar.

The method of scoring adopted both for the vocabulary and for the translation parts of the test are convenient, relatively simple, and involve a minimum of judgment on the part of the examiner.

The third desirable feature of a test is the possibility of sufficient duplication so that if pupils coach up on the test other tests of presumably corresponding value and difficulty may be substituted. This is a highly desirable feature of any educational test. When a test is being used extensively it becomes familiar to both teachers and pupils and tends thereby to lose its measuring value. This condition will have to be met by all tests that are to be used continuously and frequently and one of the important values of tests will be their continued use at frequent intervals.

The underlying motive in the preparation of the present test has been to meet this fundamental condition. The test has consequently been so constructed that it can be duplicated indefinitely with very little expenditure of time or effort. For example the vocabulary test may be duplicated to the extent of twenty sets of 100 words each which would consume the 2000 words of Lodge's list. If pupils should coach up on a given list, any other may be easily substituted and if they should coach up on the entire 2000 words so much the better. They would then have substantially the entire vocabulary of high school Latin.

The translation test can likewise be duplicated to the extent of consuming the entire text material used in the four years of high

school Latin by following the simple instructions used in the preparation of the test here presented. Later in this paper data will be presented to show the extent to which various tests made up in this manner are comparable or of similar difficulty. It will be found that they agree quite closely.

Adequacy of the test. Two questions regarding the adequacy or reliability of this test arise. First, How adequately does the present test measure the attainment in Latin? Second, How closely do the measures obtained for one test, made up in the manner here proposed, compare with any other test made up in a like manner?

Probably very little doubt concerning the general fairness and adequacy of the test will arise in the mind of anyone who makes a critical study of it for the reason that the method of the test is essentially similar to that ordinarily employed by teachers.

The second question regarding the adequacy of the method is fundamental to the viewpoint here adopted. How do various tests prepared according to the specifications previously stated compare with one another in difficulty and reliability? To answer this question definitely by means of experimental data, two sets of vocabulary and two sets of translation tests for the first year and two sets for the second year were prepared and given to the same pupils. The following table gives the scores made in the two first-year tests by a class of nine pupils in first year Latin and the scores made in the two Caesar tests by a class of twelve pupils in the second-year Latin.

Comparison of Test No. I and Test No. II in First Year Latin.
Percentage Scores.

	No. of Words.	First 10 Sentences.
Test I	122	84.3
Test II	139	83.0

Comparison of Tests No. I and No. II in Caesar.

	No. of Words.	Percentage Scores.
Test I	154	72.9
Test II	142	76.2

The scores for the two first year tests are based on the first ten sentences of the test since the pupils were not allowed sufficient time to complete all of the second test. These results show a remarkably close agreement in the measurement of the same pupils

by two different tests prepared in the same manner. There is a difference of only 1.3 points between the two first year tests and of 3.3 points between the second year tests. These differences are negligible as compared with the gross scores. It is evidently possible to prepare tests according to the principle here proposed, to use them interchangeably and to measure pupils with considerable accuracy.

A further study of the same point was made by comparing the scores made by the same pupils on the odd numbered sentences with those made on the even numbered sentences of the same test. This was done for the first year and the second year tests as shown in the following table:

Comparisons of Two Divisions of First Year Test No. I

	No. Wds. in Test	No. Pupils	Percent Correct
All of Test	122	122	68.4
Odd Sentences	57	122	69.5
Even Sentences	65	122	67.4

Comparison of Two Divisions of Caesar Test No. I

	No. Wds. in Test	No. Pupils	Percent Correct
All of Test	154	96	62.4
Odd Sentences	95	96	61.6
Even Sentences	59	96	63.5

The scores on the two halves of each test agree very closely. The difference between the two halves of the first year test was 2.1 points and between the two halves of the second year test, only 1.9 points. These comparisons corroborate the fact that measurements obtained from any one test will be very closely comparable with those obtained from any other test prepared in the same manner.

Results. The test was given in the spring of 1918 to nearly 1000 pupils in thirteen high schools in five states, namely, Illinois, Michigan, Ohio, Tennessee, and Wisconsin. The schools varied in the number of Latin pupils from twenty to 300. The scores for the different schools and the number of pupils in each school are given in the following table:

VOCABULARY TEST

Averages by Schools

School	Averages I.	II.	III.	IV.	No. of Pupils I.	II.	III.	IV.
A	45.1				16			
B	21.8	38.5		50.1	67	48		12
C	25.3	34.2		72.8	19	11		4
D	64.7		66.4	81.0	15		15	18
E	30.0	42.9		73.5	4	7		4
F	37.5	49.2	62.3		11	12	11	
G	33.7	56.8	63.3		9	13	4	
H	32.0	51.0			6	6		
I	44.4	59.5	74.0	83.3	13	28	2	3
K, L, M	29.7	39.0		64.0	(About 300)			
Ave.	30 1	45.9	67.2	69.9	160	125	32	41

TRANSLATION TEST

Percentage Scores

School	Per cent Correct I.	II.	III.	IV.	Number of Pupils I.	II.	III.	IV.
A	57.7			54.5	25			8
B	43.9	56.4		44.3	67	48		12
C	60.1	62.6		60.7	19	9		4
E	82.9	64.9		90.8	3	9		4
F	84.7	76.1	59.5		11	12	12	
G	71.6	72.9		59.9	9	12		5
H	72.3	73.4		88.6	6	6		1
J	87.3							
K, L, M	69.4				(About 300)			
Ave.	63.1	62.4	59.5	57.7	184	96	12	34

On the basis of these results the following scores may be proposed as tentative June standards (Medians):

Years of Latin	1	2	3	4
Vocabulary—No. of words	35	50	65	80
Translation—Percentage scores	70	72	62	62

The averages given for the different years are not averages of the schools but are averages of the total number of pupils who did the test.

The proposed standard scores are medians and are somewhat higher than the averages for the corresponding years owing to the fact that the averages were brought down by a small number of rather defective pupils particularly in the upper years. The differ-

ences among the scores for the different schools are exceedingly large. For example, in the vocabulary test the first year classes, ranged from 21.8 in one school to 64.7 in another school. In the translation test the first year classes ranged from a score of 43.9 to 87.3. It will be noted that school B which stood low in the vocabulary test also stood low in the translation test. Similar comparisons for the different years show similar differences. School B, one of the largest high schools in the list, is consistently low in all the years of the high school.

The next table shows the attainment at half year periods for 56 first year pupils and 184 second year pupils. The figures show the gain made in each half year.

Year I	First Sem.	Second Sem.	Difference	Percent Gain
Vocab.	14.3	30.1	15.8	110.5
Trans.	36.9	63.1	26.2	71.6
Year II	**First Sem.**	**Second Sem.**	**Difference**	**Percent Gain**
Vocab.	25.0	45.9	20.9	83.6
Trans.	40.4	63.7	23.3	57.6

The next point of interest is the distribution of the individual scores and the overlapping of the pupils in the different years of Latin.

Distribution and overlapping of pupils in the various years. The table reads thus: Thirty pupils with 1 year of Latin had a vocabulary score between 11 and 20, 49 a score between 21 and 30, etc.

VOCABULARY TEST

4th year	1	1	3	1	1	6	15	10	3
3rd year				3	14	11	5	1	
2nd year	6	18	29	25	26	17	2	2	
1st year	30	49	38	12	14	2			
Scores (Words)	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100

TRANSLATION TEST

4th year., 2nd Sem.		3	4	5	7	6	5	3	1
3rd yr., 2nd sem.				4	5	3			4
2nd sem.		6	5	7	17	21	22	12	4
2nd yr., 1st sem.	5	2	6	3	4	1	2		
2nd sem.	5	6	22	24	28	18	29	27	25
1st yr., 1st sem.	4	16	16	9	7	3	1		
Percent Scores	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100

These results are typical of what has been found in practically every school subject in which comparative tests have been applied. Thus according to the vocabulary test there are some pupils in the fourth year of Latin who know no more words than the poorest pupils in the first year; and the best pupils in the first year know as many words as the average pupil in the fourth year. A similar comparison may be made on the basis of the translation test. However, in the above table the comparison can be made only for the two semesters of a given year since a different test is used each year. A comparison of any year may be made with any other year by giving the same translation test to both classes.

The next question is the relation between the knowledge of vocabulary and the ability to translate. The following table gives the coefficients of correlation between the vocabulary and the translation test for the pupils in four classes.

School	Year	No. of Pupils	Coefficient of Correlation
B	I	18	.67
B	II	21	.82
B	IV	12	.52
C	I	19	.52

These correlations are fairly high and indicate that knowledge of vocabulary and ability to translate tend to go together rather closely.

The length of the sentence is usually considered important in translation. Longer sentences are considered much harder. The test contains sentences of varying length. The percentage of accuracy of translation for different sentences is given in the following table:

FIRST YEAR TEST

Sen. No.	1	2	3	4	5	6	7	8	9	10
Wds. in Sen.	6	6	4	5	5	8	5	4	6	12
Percent Correct	81.7	79.7	63.3	83.1	85.7	81.3	84.8	66.4	82.7	64.1
Sen. No.	11	12	13	14	15	16	17	18	19	20
Wds. In sen.	6	10	9	3	6	4	7	4	3	9
Percent Correct	71.8	61.7	58.8	57.5	73.2	65.3	44.9	58.8	50.2	57.1

This table indicates that length of sentence does not materially affect the accuracy of translation. The correlation between length of sentence and percentage of accuracy is only .02 for the first year test and .31 for the second year sentences.

Another point of interest is the comparison of the sexes. The following table gives the scores for both the vocabulary and the translation test of pupils who had studied Latin for half a year, one year, and two years. The results show no consistent difference and the probability is that boys and girls are substantially equal in their capacity to study Latin.

COMPARISON OF BOYS AND GIRLS

Years of Latin	Half Year		1 Year		2 Years	
	Boys	Girls	Boys	Girls	Boys	Girls
No. Pupils	24	32	69	86	25	57
Vocabulary	13.	14.5	25.	28.	40.	38.
Translation	29.5	35 2	59.	55	61.7	65.

LATIN TEST—SERIES A

VOCABULARY TEST

1. a, ab, abs
2. aditus
3. affero
4. alius
5. ante
6. avis
7. captivus
8. cibus
9. cohors
10. concilium
11. congredior
12. contra
13. de
14. defigo
15. detrimentum
16. distribuo
17. duplex
18. ferus
19. fossa
20. gladius
21. hospitium
22. ignosco
23. incido
24. ipse
25. iustitia
26. latus
27. maneo
28. moror
29. nemo
30. ob
31. opera
32. orior
33. passus
34. permitto
35. plerusque
36. postea
37. praefero
38. pridie
39. proficiscor
40. puer
41. quartus
42. quod
43. remitto
44. sexaginta
45. sinister
46. studium
47. supero
48. tam
49. tollo
50. triumpho
51. valles
52. video
53. auspicium
54. claudo
55. conscientia
56. curia
57. dignitas
58. erumpo
59. excito
60. exsul
61. flos
62. indicium
63. inquam
64. meus
65. municipium
66. otium
67. prospicio
68. regius
69. sanguis
70. securis
71. sollicito
72. tego
73. utinam
74. veneror
75. vito
76. aer
77. apto
78. arvum
79. caeli-cola
80. compono
81. corripio
82. donum
83. equidem
84. famulus
85. frigidus
86. gemitus
87. implico
88. laevus
89. liquidus
90. magister
91. memoro
92. nodus
93. obsto
94. pelagus
95. rapio
96. reviso
97. sacro
98. sterno
99. turba
100. uro

FIRST YEAR LATIN

1. Cur in pulchram insulam frumentum partamus?
2. Equus niger viri in silva est.
3. Agricola poetis insulam monstrat.

4. Multi milites castra Gallorum oppugnant.
5. Filia agricolae est grata dominae.
6. Caesar veteres milites amabat, quod bello fortes erant.
7. Galli et Germani dissimillimi erant.
8. Quae arma militimus fuerunt.
9. Magnum numerum equitum et peditum habebimus.
10. Ubi oppidum a perfido Sexto Occupatum est, oppidani miseri gladio interfecti sunt.
11. Legatus multo fortior est meo fratre.
12. Imperator legatum misit ut in citeriore Gallia novos milites conscriberet.
13. Ruri sunt viae pedibus multis angustiores quam in urbe.
14. In Galliam contendimus.
15. Aestate dies sunt longiores quam heime.
16. Cur hostes se receperunt?
17. Si telis uti potuissent, non victi essent.
18. Nolite consilia malorum audire.
19. Cum venies, cognosces.
20. Cum principes inter se obsides darent, Romani bellum paraverunt.

SECOND YEAR LATIN. Caesar.

1. Die constituta causae dictionis Orgetorix ad iudicium onmem suam familiam, ad hominum milia decem, undique coegit et omnes cliente obraetosques suos, quorum magnum numerum habebat, eodem conduxit; per eos ne causam diceret se eripuit.

2. Qui cum itinere convenissent seque ad pedes proiecissent suppliciterque locuti flentes pacem petissent; atque eos in eo loco quo tum essent suum adventum exspectare iussisset, paruerunt.

3. Primam et secundam aciem in armis esse, tertiam castra munire iussit.

4. His rebus cognnitis exploratores centurionesque praemittit qui locum castris idoneum deligant.

5. Brevi spatio interiecto, vix ut iis rebus quas constituissent conlocandis atque administrandis tempus daretur, hostes ex omnibus partibus signo dato decurrere, lapides gaesaque in vallum coicere.

6. Tum vero clamore ab ea parte audito nostri redintegratis viribus, quod plerumque in spe victoriae accidere consuevit, acrius impugnare coeperunt.

7. Neque enim temere praeter mercatores illo adit quisquam, neque his ipsis quicquam praeter oram maritiam atque eas regiones quae sunt contra Gallias notum est.

THIRD YEAR LATIN. Cicero.

1. Num me fefellit, Catilina non modo res tanta, tam atrox, tamque incredibilis, verum, id quod multo magis est admirandum, dies.

2. Video cui sit Apulia attributa, quis habeat Etruriam, quis agrum Picenum, quis Gallicum, quis sibi has urbanas insidias caedis atque incendiorum depoposcerit.

3. Etenim, Quirites, si ea quae erant ad me delata reperta non essent, tamen ego non arbitrabar in tantis rei publicaee periculis esse mihi nimiam diligential pertimescendam.

4. Sed ego institui referre ad vos, patres conscripti, tamquam integrum, et de factor quid iudicetis et de poena quid censeatis.

5. Verum tamen illis imperatoribus laus est tribuenda quod eregunt, venia danda quod reliquerunt, propterea quod ab eo bello Sullam in Italiam res publica, Murenam Sulla revocavit.

6. De quo homine vos, id quod maxime facit auctoritatem, tanta et tam praeclara iudicia fecistis?

7. Itaque hunc et Tarentini et Locrenses et Regini et Neapolitani civitateceterisque praemiis donuarunt, et omnes, qui aliquid de ingeniis poterant iudicare, cognitione atque hospitio dignum existimarunt.

FOURTH YEAR LATIN. Virgil.

1. Eripiunt subito nubes caelumque diemque
Teucrorum ex oculis; ponto nox incubat atra.
2. Vix ea fatus erat, cum circumfusa repente
|scindit se nubes et in aethera purgat apertum
3. Iam Deiphobi dedit ampla ruinam
Vulcano superante domus, iam proximus ardet
Ucalegon, Sigea igni freta lata relucent.
4. Vix Prima inceperat aestas
et pater Anchises dare Fatis vela iubebat.
litora cum patriae lacrimans portusque relinquo
et canoism ubi Troia fuit.
5. Sternimur optatae gremio telluris ad undam
sortiti remos passimque in litore sicco
corpora curamus; fessos sopor inrigat artus.
6. Ocius omnes imperio laeti parent et iussa facessunt.
7. Ille agmine longo
tandem inter pateras et levia pocula serpens
libavitque dapes rursusque innoxius imo
successit tumulo et depasta altaria liquit.
8. Hic primum Fortuna fidem mutata novavit.
9. At pius Aeneas ingenti mole sepulcrum
inponit suaque arma viro, remumque tubamque,
monte sub aerio, qui nunc Misenus ab illo
dicitur aeternumque tenet per saecula nomen.
10. Hinc metuunt cupiuntque, dolent gaudentque, neque auras
dispiciunt clausae tenebris et carcere caeco.[1]

---

(1) Copies of the tests may be obtained from the University Cooperative Co., Madison, Wisconsin.

DETECTING THE FEEBLE-MINDED IN A CITY SCHOOL POPULATION

Thomas H. Haines
Ohio State University, Columbus, Ohio

THE examination work reported herein was undertaken for the Board of Education of Columbus, Ohio, by the Ohio Bureau of Juvenile Research. Every school in the city was visited in the years 1915-17 and all the feeble-minded children discovered were reported to the principals of the schools and to the superintendent. The results of some of the earliest work have unfortunately been mislaid and are not now available. Some of the schools had been canvassed by students of the department of psychology of the Ohio State University. The work of these students was not duplicated since the school authorities had been informed as to the defectives in these schools. The thirty-eight schools reported herein had an aggregate enrollment of 18,223 at the time of the examinations. This represents more than three-fourths of the school population of the City.

The method of procedure in a school was for the examiner to tell the principal the object of the visit and let the principal bring forth such children as she wished to have examined. As would be expected the personal attitude of principals was very different upon this matter. In some schools the examiner was left very much in doubt as to whether all the children who might be feeble-minded were brought forward, and in other cases the tendency was to use the occasion of the psychologist's visit for getting intelligence ratings upon as many of the children as possible. The number of children selected by principals for examination bears no constant relation to the number of feeble-minded discovered. In addition to this we must bear in mind the reference of Binet to the poor quality of judgment teachers have in respect to mental abilities of their pupils, and in fact to the bias which familiarity seems to work in the judgment of any one in respect to the intelligence assessment of another.

Doubtless many feeble-minded children in the community escaped this survey by not being in school at the time of the investigation. We believe, however, that of the children attending school these

examinations have served to discover the ones so poorly endowed that they have to be classed as feeble-minded, and that we have, in the results here offered, indices of the intelligence of the different school districts of this City, as well as an approximately truthful statement of the total number of feeble-minded children amongst these 18,223.

Examination was conducted, for the most part, by the Yerkes-Bridges Point Scale method. Some small children were tested by a Year Scale; a considerable number were measured by both methods. Among the 152 children making a coefficient of mental ability of less than 67 there were only two cases that were not measured by the Yerkes-Bridges Scale. What is rated a coefficient of mental ability in each of these cases was really an intelligence quotient, as Terman uses that term. Likewise among the 244 children making coefficients of mental ability of 67 to 81, inclusive, there were five who dd not have a Point Scale measurement, and for whom we therefore report intelligence quotients. The coefficients of mental ability and the intelligence quotient are not the same mathematical quantities. Because of the very small numbers of intelligence quotients involved we have not eliminated them, or given them separate consideration.

The examinational work was done in twenty-seven schools by Miss Lillian E. Coler; in six schools by Miss Mina A. Sessions; and in five schools by Miss Alida C. Bowler. It was intended to secure the co-operation of the school physician and nurses for physical examination, and of teachers for family histories and social reactions of children. It was not found possible to secure these. The physical examination consisted of conversations with the child and the teacher in supplementing the information obtained by the test, and all such facts were carried to the office and taken into consideration in making the diagnosis. No child was classified as feeble-minded on the basis of his intelligence examinations alone.

Table I shows 157 children reported to the superintendent and principals as feeble-minded, while it exhibits 153 children making coefficients of mental ability of less than 67. We do not consider that any fixed point in the rising scale of coefficients is the point of separation between the feeble-minded and those competent to manage themselves prudently and to profit by ordinary educational facilities. We do believe that the intelligence examination is one important means of making this separation between the

TABLE I

Presenting enrollments of 38 city schools with numbers of children tested in each, numbers and rates of feeble-minded found in each school.

School	Average enrollment	Number Examined	Number making C. M. A. 67 to 80	Number making C. M. A. less than 67	Number reported to Superintendent as Feeble minded	Number per 1000 of enrollment reported as Feeble-minded
No. 1	471	20	8	6	5	10.6
No. 2	440	12	5	3	3	6.8
No. 3	416	48	22	6	5	12.0
No. 4	423	13	20	8	7	16.5
No. 5	437	14	4	2	2	4.6
No. 6	530	6	3	2	1	1.9
No. 7	259	7	4	1	1	3.9
No. 8	482	19	6	3	0	0
No. 9	500	15	10	4	4	8.0
No. 10	426	34	10	12	9	21.2
No. 11	550	32	14	4	4	7.3
No. 12	591	16	2	2	2	3.4
No. 13	418	26	5	11	8	19.1
No. 14	558	7	2	4	3	5.4
No. 15	559	8	0	2	2	3.6
No. 16	673	17	6	5	4	5.9
No. 17	541	16	4	1	2	3.7
No. 18	746	25	6	4	5	6.7
No. 19	526	19	4	1	1	1.9
No. 20	616	9	3	2	2	3.2
No. 21	341	48	16	8	11	32.3
No. 22	634	22	13	4	7	11.0
No. 23	415	34	6	1	1	2.4
No. 24	324	18	3	11	11	33.9
No. 25	464	20	6	5	6	12.9
No. 26	470	14	4	2	3	6.5
No. 27	141	22	8	4	6	42.6
No. 28	384	7	2	1	2	5.2
No. 29	393	13	0	0	0	0
No. 30	448	36	10	12	14*	31.3
No. 31	514	8	1	1	1	1.9
No. 32	716	13	3	1	2	2.8
No. 33	641	15	4	6	6	9.4
No. 34	458	12	3	2	3	6.5
No. 35	332	12	1	0	0	0
No. 36	446	23	6	8	9	20.2
No. 37	112	16	2	4	5	44.6
No. 38	828	11	0	0	0	0
Totals	18,223	707	226	153	157	8.6

*5 children found feeble-minded by Ohio State University not included

feeble-minded and those who are competent, and the coefficient of mental ability of 67 is about the point where, as we descend in the scale of intelligence, doubt as to mental competence enters. Some children who make coefficients of mental ability of more than 67 must be rated as feeble-minded from the evidence furnished, from moral attitudes, educational, vocational and social facts, or from physical defects. Others who make coefficients of mental ability of less than 67 for similar reasons must be diagnosed as competent to manage themselves prudently and without offense to other members of the community.

Comparison of the numbers reported feeble-minded and the numbers making coefficients of mental ability of less than 67, school by school, exhibits our practice. School No. 10 shows twelve children making a coefficient of mental ability of less than 67, and only nine children were reported feeble-minded; and school No. 21 shows eight children making a coefficient of mental ability of less than 67 and eleven reported feeble-minded. Careful conservatism was exercised in every case. No child was reported to the superintendent and principals as being feeble-minded where any reasonable ground for question or doubt existed.

It will be noticed by observing the totals at the bottom of Table I that 157 children out of 18,223 enrolled, or at the rate of 8.6 per thousand were reported feeble-minded.

The last column of this table giving the numbers per thousand of enrollment found to be feeble-minded, school by school, exhibits most striking variations in prevalence of mental defect in different school districts. There are four schools with an aggregate enrollment of 2040 children in which no feeble-minded child was discovered. Two other schools, No. 27 and No. 37, with 253 enrolled have eleven feeble-minded children, or about 43 per thousand, or 4.3 per cent of the population. Three other schools, No. 21, No. 24, and No. 30, with an aggregate enrollment of 1113 have over 30 feeble-minded children per thousand.

In Table II, the 396 children with coefficients of mental ability ranging from 12 to 81 are distributed according to ages at their last birthdays, and according to the coefficient of mental ability. Of these, 152 were below 66 and 244 from 67 to 81. It is to be observed that some very defective children remain in school beyond the fifteenth year of age, but the maximum number of defectives and near defectives of this group is passed at the tenth birthday. The largest number of children of a given year of age making coefficients

TABLE II

Distribution according to *C. M. A.* and *age* of 396 children making C. M. A. 12 to 81

Ages		Coefficients of Mental Ability											Totals by Ages	Coef. of M. A.			Totals by Ages	Total
		12 to 16	7 to 21	22 to 26	27 to 31	32 to 36	37 to 41	42 to 46	47 to 51	52 to 56	57 to 61	62 to 66	12 to 66	67 to 71	72 to 76	77 to 81	67 to 81	12 to 18
6	years		1									1	2	4	4	2	10	12
7	years							1	3	3	4	2	13	5	8	5	18	31
8	years	1			1				2	2	2	7	15	2	6	4	12	27
9	years				1		4	1	5	2	6	5	24	2	10	14	26	50
10	years			1	1				2	1	7	7	19	7	9	12	28	47
11	years					3			1	7	3	3	17	15	9	10	34	51
12	years				1		2	2			5	8	18	15	10	13	38	56
13	years						1		3	4	3	5	16	6	9	12	27	43
14	years					1	1		1	3	5	4	15	6	7	12	25	40
15	years					1				2	3	6	12	5	10	3	18	30
16	years											1	1	2	3		5	6
17	years													1		2	3	3
Totals		1	1	1	4	5	8	4	17	24	38	49	152	70	85	89	244	396

of mental ability of less than 67 is at nine years where there are 24 children. Still it will be observed that out of the 152 making these low coefficients of mental ability 98, or nearly two-thirds of the whole number, are more than nine years of age. Large numbers of feeble-minded children are allowed to clog the machinery of a regular school and impair the progress of the normal child. This is particularly true in the tenth, eleventh, and twelfth years of age.

TABLE III

Distribution according to *C. M. A.* and *School Grade* of 396 children making C. M. A's 12 to 81.

School Grade	Coefficients of Mental Ability											Totals C. M. A.				Totals	
	12 to 16	17 to 21	22 to 26	27 to 31	32 to 36	37 to 41	42 to 46	47 to 51	52 to 56	57 to 61	62 to 66	12 to 71	67 to 76	72 to 81	77 to 81	67 to 81	12 to 81
Special					1				1	2	4	8	3	4	9	16	24
I	1	1	1	3	2	4	3	9	7	9	10	50	11	16	8	35	85
II				1		1	1	3	4	13	10	33	6	14	14	34	67
III						3		2	5	6	9	25	14	13	11	38	63
IV					1			3	5	3	7	19	19	14	24	57	76
V					1				2	4	7	14	11	14	9	34	48
VI										1	1	2	5	8	9	22	24
VII											1	1	1	1	4	6	7
VIII														1	1	2	2
	1	1	1	4	5	8	4	17	24	38	49	152	70	85	89	244	396

In Table III the children making the same coefficients of mental ability (396 children) are distributed by the grades in school making them at the times of examination and by the coefficients of mental ability. The maximum number in single year of school is found in the second grade where there are 50 with coefficients of less than 67, and 35 with coefficients between 67 and 81. This is also de cidedly the maximum number of the former group, the next largest number being found in the third grade (33). Both morons and dullards *accumulate* in grade two, and both groups clog the machinery of grades three to five. Relatively small numbers of feeble-minded children are allowed to pass grade four.

TABLE IV

Distribution of the 152 children making C. M. A's of less than 67 by *age* and school grade.

	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.	11 yrs.	12 yrs.	13 yrs.	14 yrs.	15 yrs.	16 yrs.	Totals
Special					1		1	2	1	3		8
Gr. I	2	12	12	14	4	4	1		1			50
Gr. II		1	3	9	8	4	6	2				33
Gr. III				1	6	6	7	2	3			25
Gr. IV						2	1	8	4	4		19
Gr. V						1	2	2	6	3		14
Gr. VI										2		2
Gr. VII											1	1
Totals	2	13	15	24	19	17	18	16	15	12	1	152

Table IV presents the distribution by age and grade of the 152 children making coefficients of mental ability of less than 67. This Table exhibits the slow resort to the special school and special room for even the feeble-minded child, and it exhibits the omnipresent tendency to push the feeble-minded on to higher grades where the seats will accommodate their growing bodies, even though they cannot profit by their experiences in the advanced room. In fact an indefinite hope that the child can find some means of growth in some other teacher mingles, in the mind of principal and teacher, with the desire to be rid of the *impediment* which the feeble-minded child constitutes. It is the school's business to prepare the child for promotion and to promote him. The feeble-minded child for whose socialization the mechanism of the school is poorly arranged falls a victim to the machine and he is promoted mechanically, just as if he were socialized and ready to move on.

These feeble-minded children are from one to six years behind the grades expected for their ages. There is observable, however, even allowing for the talents they do possess, or rather their distinct lack of talent, a positive tendency to advance them in school work, in which work they are making no real progress. To any one who is familiar with the eight year old mind trifling with the school room activities of the fifth grade, the facts presented in Table IV indicate a striking misapplication of educational forces in the schools of this City. Such a person realizes what a draw-back the presence of these 152 children is in these schools; how they hinder the bright

boy and girl by taking undue energy of the teacher, all to no purpose; how they fatigue and exasperate the teacher; and how they themselves fret and reap no benefit from the procedure. Such persons know that the presence of these defective children, these non-potential citizens in the public schools is a positive waste of social substance. This waste is just as real as that of the food profiteer who allows a cargo of potatoes to rot on the pier in order to keep prices up.

Such an intelligence survey of a City school population is most valuable in indicating the different levels of intelligence of various parts of the population of the City. It affords indispensable information for the intelligent planning of the education, or preparation for citizenship, of the different parts of a city population. Educational work must be adapted to the needs of the individual. We have not given so much attention to the necessity of adapting the work of character building to different wards and precincts of a city. We must take the situation as it exists in the group for whose service the school is provided in order to make citizens of the children of that group. Figures from different school districts such as this study supplies afford one of the necessary bases upon which such intelligent planning of educational work, school, by school, must be founded. All social upbuilding will find similar profit in consulting such data. Doubtless the applications of psychology to community service will be greatly broadened in the near future. We shall have more elaborate means of analyzing the potentialities of citizenship in the individual. For the present, however, this is no argument for neglecting the very useful information which applied psychology is placing in the hands of the community builder. Such figures on intelligence measurements are an indispensable part of any school survey which looks to the more fruitful application of educational forces to character building in a community and to the development of a better democracy.

THE RANGE OF INFORMATION IN BIOLOGY III. BOTANY*

N. M. Grier, Ph. D.
Hollins College, Hollins, Virginia

I. Introductory

IN the preceeding paper of this series (1), it was pointed out that the most closely related predecessor of secondary biology was probably the nature study of the grades, since it seems that the larger portion of nature study as taught at present consists of extremely elementary zoology and botany. The viewpoints disclosed by the range of information test in the courses of physiology and zoology have been discussed in the two previous papers, (1-2), and of the components of nature study which we have recognized, there remains only botany to treat in a similar manner.

Further insight as to the amount of botanical knowledge pupils carry with them into the high schools seems desirable for a number of reasons. Botany in the first place shares with the other secondary biological sciences the condemnation implied by a decline in the number of students taking them in recent years, as pointed out by Downing (3). Leading botanists are inclined to attribute this situation to the want of effective teaching (4). Under ordinary circumstances, one could consistently take the standpoint that, since no form of life is more abundant than plants and none so important, there should be no insuperable difficulty in making the teaching of such a science effective. That there has been some comprehension of this advantage is shown by the variety of textbooks on hand, apparently adapted to every type of locality in which botany is given. Textbooks dealing with some applied form, such as agriculture, are available for rural districts where they are more likely to be desired; for cities and larger towns, where often more purely cultural values seem to be rated higher, there are others colored with the philosophical aspects of botany, such as plant ecology.

---

*The writer is indebted to the authorities of the St. Louis Schools, Prof. J. A. Drushel, Harris Teacher's College, Dr. G. M. Holferty, Central High School, St. Louis, for aid in furthering this investigation.

To the best of the writer's knowledge no textbook of botany in use at this time takes into serious consideration any previous knowledge of botany gained from the nature study of the grades. There may be adequate reason for this. Nature study is to a great extent synonomous with natural history, but so far as instruction goes the natural history of plants seems on the whole comparatively neglected when compared with that of animals, as parts of it, such as the complicated life cycles of certain groups of plants, seem too difficult for pupils of an early age. It would be fair to assume that botany has gravitated to its present place in the older type of high school on account of being more closely related to the interests of the pupils at least, if not on account of any relation with nature study. The course in nature study, however, constantly tends to include more botanical knowledge within its scope, and the existing bond between it and botany is being strengthened. Those interested in the effective teaching of high school botany should take account of this fact, and may well inquire to what extent the ground has been broken by nature study. Such has been the effort of the writer based on the range of information test (5).

II. Method

The 100 test words were selected in the following manner: A list was first prepared of all the words pertaining directly to botany in the spellers and readers previously used by the pupils in their grade school work. A similar list was made from Boy Scout and Camp Fire girls publications. From these lists, 100 of the more frequently occurring words were checked, those chosen being:

1. Words in fairly common use, a knowledge of which would indicate the range of information of those phases of plant life the pupils might be expected to know about from their nature study, school gardens, etc.
2. Words, a knowledge of which would indicate some familiarity with biological principles, and the relation of plants to lower animal groups as well as to man.

Since the method of giving the test has been discussed in previous papers there is given here only the list of words used. It need only be said that in giving the test the pupils were told how the terms were selected and that they pertained directly or indirectly to plants.

III. Results

The results are found in the following tables. By first term students is meant those taking up the study of botany at the time. Second term students have just finished botany and are beginning physiology. It may be well to remark that as a check on the original marking of the pupils, (first series), the 100 words were defined from the glossaries of textbooks containing the words, and each pupil revised his own paper by placing a "second series" of marks after each word, indicating the manner in which he should have marked it. A comparison of the D's, E's, F's, N's, of the first and second series, besides showing fairly accurately the extent and nature of the error due to ignorance or misunderstanding of the real meaning, gives a rough preliminary measure of the extent to which these 100 terms are understood by first year high school pupils.

TEST WORDS

absorption	crown	mould	seed leaf
alternate bud	desert	nectar	seedling
alternate branch	digestion	nitrogen	sepal
alternate flower	energy	nut	sheet
animal	evergreen	opposite bud	shrub
anther	fermentation	opposite branch	spines
bacteria	fern	opposite flower	spray
bark	fertilizer	oxygen	stalk
berry	fibers	parasite	stamen
biennial	fibrous root	perennial	starch
blade	floret	petal	stem
blight	flower	pistil	stigma
branch	foliage	pith	swamp
bud	fruit	plantation	tap-root
bulb	fungus	pod	tassel
cactus	gall	pollen	tendril
calyx	germinate	proteid	terminal bud
carbohydrate	grain	ray	trunk
carbonic acid gas	herb	resin	tuber
cell	leaf	reproduction	twig
coal	leaflet	ripe	vegetable
composite	legume	rootlet	veins
corolla	meadow	rosette	vine
cork	migration	sap	woods
cross	moss	seed coat	yeast

TABLE I

DEPENDENCE OF RANGE OF INFORMATION ON ACADEMIC STATUS

A

(1st series)

Academic status	No.	D	E	F	N	%Error
1st term pupils	97	22.86	19.67	19.99	37.07	
2nd term pupils	86	32.4	31.70	20.70	15.05	

B

(2nd series)

Academic status	No.	D	E	F	N	%Error
1st term pupils	97	26.84	25.25	20.05	26.64	4.7
2nd term pupils	86	39.96	31.72	16.32	11.64	3.84

Table I, A & B, shows that with maturity there is a decided increase in the number of terms that can be defined, (D), explained, (E), with a corresponding decrease in the number of new terms, (N). In the preceding two papers of the series, the greatest increase was found to be in the number of F's. Here they are practically equivalent in the first series, and are found to diminish in the second. In both grades of pupils, checking up as a whole increases only the D's, and diminishes only the N's; the E's are increased only in the first term pupils, whereas in the preceding two papers it was found that the E's increased and the F's decreased in both.

A very short period of time elapsed between the completion of the botany course by the second termers and the taking of the test. While studying botany they came in contact with most but not all of the test words. This of course would explain their marked superiority, greater accuracy of definitions in D's and E's, both series. The botanical part of nature study courses is, of course, not as highly organized as the botany courses of the high schools, and the difference between the number of D's and E's of the 1st and 2nd termers is illustrative of the difference between pursuing the regular, organized study of a subject and pursuing it sporadically without constant supervision. Completion of the former type of course may be inferred to leave an ability to define or explain more words pertaining to it, while to a student of the latter type, there are more words which are merely familiar or even new. In

the first type of study ability to define exceeds ability to explain (as shown in the 2nd series), while in the second type of course these abilities are practically equal.

That the E's have not increased in the second series is probably accidental. It is likely that the same words were marked similarily in both series on both sides, but more probable that some of the E's of the first series have been displaced into D's in the second at the eventual expense of the F's and N's. There is thus a shifting forward of the values of D's and E's. The same situation is found in F's of the 1st termers and may be similarily explained. From this one may conclude that to the student untrained in botany 20% of the representative terms are generally familiar, while the trained student may explain 31% rather than merely know them. Discriminatory power in students has advanced as shown by the percentage of error, 4.7% in the first termers, and 3.84% in second termers. This of course is due to the more constant supervision of organized class room work, especially trained teachers, etc.

The D's were more numerous and the N's were fewer in the botany test than in either of the preceding tests, physiology (1) and zoology (2), possibly because illustrative material is more abundant in botany. This may also account for the lowest percentage of error in it of all three subjects, and may be taken as an indication that *high school botany as a subject has the best basis of all the biological subjects upon which to build, and probably more elementary botany could safely be included in the course of nature study. With such an aid to effective teaching there would seem to be no excuse for its failure to persist in the curriculum.*

TABLE II

DEPENDENCE OF RANGE OF INFORMATION ON SEX

A

(1st series)

Academic status	No.	D	E	F	N	%Error
High School Boys	101	31.88	24.18	18.42	25.03	
High School Girls	82	23.38	27.2	22.28	27.39	

B

(2nd series)

Academic status	No.	D	E	F	N	%Error
High School Boys	101	34.27	26.49	16.68	21.95	7%
High School Girls	82	32.58	30.48	19.79	17.42	6%

Reclassified by sex, the results of the rest show a slight superiority in girls of range of information in nature study. This somewhat substantiates the old tradition of botany being a "girls' school subject." Another conclusion derived is the same as that taken from the paper on zoology: That these particular concepts are clearer with boys than girls, regardless of what boys may retain. This checks with the percentage of error, 2 per cent with boys, 7 per cent with girls. It was previously found that the boys showed a greater range of information in physiology, but it seems that the girls excel in nature study as shown by this and the paper on zoology. The percentage of error is practically the same as in the latter paper, 2% in boys, 6% in girls, the same as that obtained in both parts of the succeeding tables. When these are compared with physiology —5% and 8%, one may infer that the latter is a more difficult subject for pupils.

TABLE III

A

(1st series)

Academic status	No.	D	E	F	N	%Error
2nd term High School Boys	54	42.86	29.48	14.29	12.67	2%
2nd term High School Girls	32	25.4	33.3	24.2	17.1	7%

B

(2nd series)

Academic Status	No.	D	E	F	N	%Error
2nd term High School Boys	54	42.86	29.48	14.29	12.67	2%
2nd term Hgh School Girls	32	37.06	33.96	18.34	10.61	7%

The insight afforded by a mixed classification by sex and academic status is presented in Table III, A. B. C. D. Second term boys secured a greater percentage of D's in both series. Girls show a great number of E's and F's and in the second series a small number of N's. The percentage of errors in girls was 6%, in boys 2%. The results differ from that secured in zoology, in that the second term boys know more than girls and in addition have clearer ideas.

TABLE III

DEPENDENCE OF RANGE OF INFORMATION ON SEX AND ACADEMIC STATUS

(1st series)

Academic status	No.	D	E	F	N	% Error
1st term High School Boys	47	24.36	18.25	19.63	37.06	
1st term High School Girls	50	21.37	21.1	20.36	37.68	

D
(2nd series)

Academic status	No.	D	E	F	N	%Error
1st term High School Boys	47	25.68	23.5	19.10	31.23	
1st term High School Girls	50	28.	27.	21.23	24.23	

The tables dealing with first term boys and girls show that the range of information is practically equivalent in both in the first series, while the girls are generally superior in the second. Girls are superior throughout in E's and F's; they were found superior in E's in zoology, but with a greater percentage of error. To offset this apparent lack of range of information, the boys show less percentage of error, 2% against 6%.

A list of typical errors of definition follows in which the words that are apparently the source of confusion are indicated in parenthesis. The list as formulated does not have the wide range of those given in the preceding papers, possibly because as already stated botanical objects are the more familiar ones and are therefore more easily defined or explained.

Absorption—to drink; taking in bread and water; evaporating; to dry up.
Alternate bud—a bud that keeps in touch with another bud running together.
Bacteria—little animals; germs found in plants and animals.
Bark—material a tree is made of.
Berry—fruit of certain trees; a little bud that grows on a bush.
Biennial—a plant which blooms twice a year, (semi-annual); half yearly.
Blade—part of a plant.
Blight—to kill a plant by cold or frost.
Bud—beginning of a tree or flower; a baby flower before opening; an undeveloped plant.
Bulb—part of a flower; a root leaves spring from; a pod which contains flowers and roots; a plant which gets all its nourishment from water.
Cactus—grows in a desert.
Carbohydrate—a fat for a body; given off from the body carbon dioxide.
Cell—place where seeds are found; a bacteria; small place in a flower.
Fern—a flower which does not have buds.
Grain—fruit which we can eat.
Petal—part of a tree.
Pith—kind of powder inside a twig.
Shrub—a low flower.

REFERENCES

(1) GRIER, N. M. *The Range of Information Test in Biology. I Physioogy..* Journal Ed. Psychology, Vol. 9, 1918; 210-216.

(2) GRIER, N. M. *The Range of Information Test in Biology. II Zoology.* Journal Ed. Psychology, Vol. 9, 1918. 388-392.

(3) DOWNING, E. R. *Enrollment in Science in the High Schools.* Science, N. S. Vol. 46, 1917. 351-52.

(4) GAGER, C. S. *The Near Future of Botany in America.* Science, N. S. Vol. 47, 1918. 101-115.

(5) WHIPPLE, G. M. *A Range of Information Test.* Psych. Review 16, 1909. 347-351.

COMMUNICATIONS AND DISCUSSIONS

SUGGESTIONS WITH REGARD TO PROFESSOR THURSTONE'S "METHOD OF CRITICAL SCORES"

THE writer was much interested in Professor Thurstone's vigorous demand, in a recent article in this Journal [1] for a greater consideration of the practical problems of individual diagnosis, in evaluating mental tests. The few comments below are intended not as criticism of Professor Thurstone's methods, but rather as an effort to carry his general plans still further than he has done.

In the first place, if a group of tests is to be used for the purpose of prognosticating failure in college why not evaluate the tests directly according to their ability to do this, – – instead of by the more round-a-bout comparison of their agreement with instructors estimates of ability? Various rough and ready methods for measuring the "efficiency" of a test in differentiating the extremes of a group are possible. The writer has recently been working over some results from a large sixth grade class in Junior High School, with a brief group scale of intelligence, in order to compare the differential value of the tests of this scale in distinguishing those pupils who failed at the end of the school year. The method used is extremely simple and straightforward. The results for the sixth grade on each test had been already tabulated. At the end of the year the writer simply obtained from the school a list of the failures and marked the score of each one of these children over in red ink. This, in itself, gave a rough picture of the situation; the extent to which the red marks of the failures were found well up among the pencil marks indicating the scores of those who passed in their work for the year showed the extent to which the particular test in question failed to differentiate the failures from the successful children. A numerical statement of differential power was obtained simply by counting the number of pencil marks below each red mark, – – that is, the number of successful children which the test failed to distinguish from each particular failure – – and averaging the results for all the red marks, for each test. This final average gave the "average error" which, expressed in percent, gave a "percent of error" for each test. The method has been described more fully elsewhere [2], and need not be further detailed here.[3] As illustrative

---

[1] Thurstone, L. L. Tests for College Freshmen, *Journal of Educational Psychology.* Vol. 10, pp. 129-92, March 1919.

[2] Pressey, S. L. and L. W. The Practical "Efficiency" of a Group Scale of Intelligence, *Journal of Applied Psychology.* Vol. 2, pp. 68-80, March 1919. A preliminary description of the "Cross-out" scale, from study of which the illustrative material of the present brief paper has been drawn, appeared in the June number of the same journal; a final statement will be published shortly.

[3] The further point should be mentioned, however, that although the method weights, heavily, marked discrepancies between test finding and the independent criterion it does not, as does the "method of critical scores" put the situation altogether at the mercy of the single extreme deviate score.

of the general run of the figures to be expected it may be said shortly that the writer's tests have shown from four to twelve per cent error in differentiating failures.

(2) A second suggestion has to do with the form that the recommendation to the school authorities, based on the test findings, should take. Professor Thurstone tells us that "Those students are reported as doubtful who attain a median percentile rank of 22% or below, and who are below the lower critical score on two or more of the six tests" (p. 136). But an administrative officer would surely be interested to know the degree of doubtfulness implied in this statement. He might also wish to know what inference might be drawn from a score at the five percentile or at the twenty-five percentile. A method which gives an exact statement of the probability of failure implied by a score at any point in the distribution may be illustrated by further treatment of the results from the "Cross-Out" scale above mentioned; results are again from the sixth grade, (156 cases). The tests were given at the beginning of a semester: at the end a list of the failures was obtained, as stated above, and also a list of children who were considered by at least three out of the four teachers in the four promotion subjects, to be doing brilliant work. There were 14 children doing exceptional work and 13 complete failures (children who failed in four or more subjects). The scores of the failures were marked in on the distribution of the total scores for the entire grade in red ink; the scores of the brilliant children in black ink. The deciles for the whole distribution were now calculated, and the number of failures and of brilliant children scoring in each tenth of the distribution noted and expressed as a percent of the total number in the division. The final result was the following table of probabilities:

TABLE I

Decile divisions	1	2	3	4	5	6	7	8	9	10
Percentage Probability that case will fail	47	13	20	7	7	0	0	0	0	0
Percentage probability that case will do brilliant work	0	0	0	0	0	0	7	20	20	47

For use in individual diagnosis such a table should be read: if a given child in the sixth grade, at the first of the semester, scores in the lowest tenth of his class, there is a 47% probability that he will be a complete failure at the end of the year; if he scores in the second decile division, there is a 13% probability etc. for the other divisions. The second line is read similarly for prognosis of brilliant work. The above figures are of no great importance as they deal with prognosis based on a relatively few cases, over a short period of time, but the method would appear to have some general application.

(3) Professor Thurstone very correctly points out the inadequacy of the correlation coefficient in giving information with regard to the ability of a test to differentiate extreme cases. However, the problems of individual diagnosis may also arise with regard to cases around the average in ability, – – as in making up "fast," "medium" and "slow" sections in Junior High School. A way in which a scatter diagram may be thrown into a simple form readily interpretable by psychological laymen, for purposes of individual prognosis in such situations, is given below. The basis of the calculations was a correlation table for the sixth grade, showing the relationship between test scores and pooled ratings of each pupil's general ability obtained, by use of a rating scale, from four teachers. The ratings were made in June after the teachers had had a year of acquaintance with the children. This table was divided into 25 parts by drawing lines through the 20, 40, 60, and 80 percentiles on both distributions; the table was thus divided each way into five sections, not equal in size on the plotting paper, but equal in having the same number of cases. It was then summarized to read for 100 cases, twenty in each of the five vertical, and twenty in each of the five horizontal divisions. The results are given in table II; 1 is high and 5 low in each distribution.

TABLE II

Division into fifths according to test scores	Division into fifths according to teachers' estimates:— 5	4	3	2	1
1	0	1	1	6	12
2	2	3	4	5	6
3	3	5	6	5	1
4	5	5	6	3	1
5	10	6	3	1	0

The table is to be read as follows(supposing the scale to have been given to a sixth grade at the beginning of a semester for the purpose of obtaining a prognostication as to ability each child will have shown at the end of the semester). If a given child scores below the 20 percentile, the probability is 10 out of 20 that he will be considered by the teachers, after a year's acquaintance, to belong in the poorest fifth of the class as regards ability; there are 6 chances out of 20 that he will be considered in the second fifth; there is only one chance out of 20 that he will be considered above the third, or average, group. If the pupil tests as in the middle fifth of the class, the chances are 3 out of 20 that he will be judged in the poorest fifth, and one out of 20 that he will be judged in the best fifth, – – and so on.

As methods of presenting test data for use in individual diagnosis these two tables have the practical advantages that (1) they are readily understood, (2) the practical implication of the test findings is direct and unmistakeable, (3) an exact statement of the definiteness of that implication, for any case, is given.

SIDNEY L. PRESSEY
Department of Psychology, Indiana University.

[4] It has been outlined in a somewhat different way in a previous article. Irregularity on a Psychological Examination as a Measure of Mental Deterioration: *Journal of Abnormal Psychology*. December, 1918.

THE RELATION OF SCHOLARSHIPS TO THE YERKES AND TERMAN ADULT TESTS

During the present collegiate year, students taking courses in Psychology and Education were tested by both the Yerkes-Rossy* adolescent-adult point scale and the Terman adult tests, by the group method. No freshmen are admitted to these classes. The results have been tabulated for the other three classes separately and for the Juniors and Seniors combined as the two latter classes take very nearly the same collegiate work. An Extension class of teachers, the majority of whom had taken 2-4 years of college work and in addition had much experience in teaching, were also given these tests. Two small classes of high school Juniors and Seniors took the Yerkes-Rossy tests. Thus there was here opportunity to compare scholastic performance with intelligence rating by two scales and also note possible diagnostic values of the different exercises.

The following tables give the mean and mean variation of each of these groups for each of the twenty exercises of the Yerkes-Rossy scale.

*Yerkes and Rossy, A Point Scale For the Measurement of Intelligence, Boston Medical and Surgical Journal CLXXVII, 564-573.

Test	50 Sophomores Mean	M. V.	50 Seniors Mean	M. V.	81 Jun.-Sen. Mean	M. V.	131 Students Mean	M. V.	17 Teachers Mean	M. V.	Total 148 Mean	High School Mean	M. V
1	6.80	1.1	7.16	1.03	7.10	1.58	7.00	1.07	7.60	.59	7.06	6.15	.68
2	3.0	0.0	3.0	0.0	3.0	0.0	3.0	0.0	3.00	00	3.0	2.84	.15
3	3.16	.86	3.56	.99	3.57	.85	3.53	.85	3.40	.97	3.52	3.02	.80
4	3.34	1.04	3.90	.66	3.90	.72	3.70	.84	4.30	.43	3.77	3.13	1.01
5	3.96	.74	3.94	.56	3.90	.59	3.91	.64	3.80	.45	3.90	3.20	.68
6	4.88	1.01	5.30	.80	5.27	.82	5.11	.88	5.70	.43	5.20	4.59	1.18
7	4.28	.56	4.38	1.16	4.45	.61	4.39	.60	4.64	.46	4.42	3.72	.85
8	6.12	1.20	6.50	1.15	6.43	1.00	6.31	1.08	7.30	.77	6.43	5.43	1.40
9	4.19	.18	4.90	.18	4.91	.15	4.91	.16	4.90	.19	4.91	4.56	.56
10	4.80	1.54	5.54	.80	5.23	.88	5.07	.77	6.00	0.0	5.18	4.61	1.00
11	3.68	.48	3.80	.33	3.84	.20	3.76	.31	3.80	.18	3.79	3.66	.45
12	5.44	1.57	5.86	1.10	5.94	.64	5.74	1.26	7.53	1.09	5.95	5.15	1.62
13	3.22	.84	3.42	.76	3.47	.26	3.37	.80	3.76	.47	3.54	2.33	1.28
14	1.86	1.04	3.72	2.45	3.80	2.48	3.06'	2.44	5.20	2.06	3.31	1.00	1.48
15	1.54	.58	1.28	.69	1.38	35	1.44	.64	1.47	.56	1.44	1.05	.72
16	3.48	2.10	3.88	2.05	3.74	2.09	3.64	2.07	4.30	1.60	3.71	2.90	1.97
17	1.52	.67	1.68	.51	1.74	.41	1.66	.52	.47	.72	1.52	1.67	.51
18	1.34	1.42	1.58	1.39	1.50	1.40	1.43	1.44	3.41	.90	1.66	1.23	1.39
19	1.80	.32	1.84	.26	1.84	.27	1.82	.29	1.82	.29	1.82	1.95	.09
20	3.60	.59	3.65	.55	3.50	.68	3.54	.65	3.94	.11	3.59	3.71	.45
Total	72.82	6.87	78.76	6.06	78.70	6.44	76.42	6.16	87.50	3.88	77.72	66.23	6.71

It is apparent from these results that this group of 50 college seniors obtained lower averages than those of the college women reported by Yerkes and Burtt in exercises 5, 7, 12, 13, 16, 18, and somewhat higher in exercises 1, 4, 8, 14, 17. The averages of all these 148 college women are lower than those given by Yerkes and Burtt for tests, 5, 10, 12 13, 16, 18, and a little higher for tests 1, 6, 7, 8, where some of the discrepancy may be due to differences in scoring. The total averages are on the whole very similar, tho somewhat higher for these seniors. It is interesting, if not significant that the 17 teachers of the Extension class exceeded the men's scores for tests 14 and 18, as well as their general average.

Another interesting comparison with the results of Yerkes and Burtt* is in the distribution of point-scale scores, as can be seenfrom the following table which gives the percentages of each group.

TABLE II

Scores	50 Seniors	31 Juniors	50 Sophomores	17 Teachers	148 Total	39 High Sch. Jun.-Sen.
40-49	......	......	2.0%	......	.6%	2.5%
50-54	......	......	......	......	0.0%	7.7%
55-59	......	3.0%	4.0%	......	3.3%	10.2%
60-64	4.0%	3.2%	14.0%	......	6.7%	25.1%
65-69	8.0%	6.4%	8.0%	......	6.7%	7.7%
70-74	14.0%	16.1%	22.0%	......	15.5%	33.3%
75-79	28.0%	25.8%	32.0%	11.7%	27.0%	10.2%
80-84	20.0%	19.6%	16.0%	5.8%	16.9%	2.5%
85-89	20.0%	16.1%	0.0%	41.1%	14.8%	......
90-94	6.0%	9.7%	2.0%	35.3%	8.8%	......
95-100	......	.....	......	5.8%	.6%	......

*Yerkes and Burtt, Relation of Point Scale Measurements to Educational Performance, School and Society, V, 535-540.

The percentages of the upper scores are larger than those obtained for the women at Harvard, and the distribution of the entire college group of 148 women compares favorably with that obtained for their men, except at the lower end. The one student whose score fell below 50 has been asked not to return on account of poor scholarship. The conspicuous feature of the distribution of these scores is that it conforms so regularly to the expected distribution of scholarship of the different groups.

In order to estimate the diagnostic value of each of the twenty exercises the percentages of the maximum score of each one obtained by the high school students were compared with those of the college seniors and of all the college students. According to this comparison, the scores of the high school girls were equal to or slightly better than those of the college students in tests 2, 11, 17, 19, 20 and 7% lower in exercise 9, and from 10% to 15% lower in tests 1, 3, 4, 5, 6, 7, 8, 10, 12, 15, 16, 18, while in tests 13 and 14 they were 30% and 37% lower respectively. A similar comparison between the 50 college seniors and the 50 sophomores showed that the scores of the former averaged 5% to 10% higher in exercises 4, 6, 10, 12, 13, 16, 17, 18, in test 14 over 26% higher, but in test 15, 13% lower, and that the average scores for the other exercises were almost equal. Hence, tests 2, 11, 19, and 20 seem of doubtful value for differentiating groups of the kind here examined.

With a few exceptions, due to changes in classes, the same college students took the Terman tests for years XIV, XVI, XVIII, and the teachers in the Extension class were given all the adult tests. The percentage of each of these groups that passed each of the tests is shown in the following table:

TABLE III

Terman Tests	XII	XIV					XVI					XVIII					
	3	2	3	4	5	6	2	3	4	5	6	1	2	3	4	5	6
48 Seniors	85.4	77.0	91.6	83.3	87.5	83.3	93.7	89.5	95.8	89.5	68.7	1.00	58.3	60.4	83.3	58.3	45.8
44 Juniors	81.8	68.1	86.3	81.8	84.0	59.0	97.7	81.8	79.5	77.2	70.6	1.00	50.0	63.6	90.9	65.9	68.1
48 Sophomores	77.0	79.1	85.4	87.5	87.5	83.3	97.9	89.5	85.4	85.4	85.4	1.00	64.5	31.2	87.5	52.0	22.9
17 Teachers	29.4						1.00	1.00	1.00	94.1	1.00	1.00	1.00	58.8	88.2	58.8	94.0
140 Students	81.4	7.50	87.8	84.3	86.4	75.7	96.4	87.1	87.1	84.3	75.0	1.00	57.8	51.4	87.1	58.5	45.1

It is evident from these results, that for these teachers the Ball-Field test was much more difficult than any of the others, and that the repetition of the digits was a much harder task than the Binet Paper-Cutting and the Ingenuity tests. The scores of the college students show that with exception of the vocabulary and Logical Memory tests, the XVIII year exercises are the most difficult, tho the clock problem is the next hardest test for the Juniors. These students found the Induction, the Code, and the Clock problems second in difficulty to the four of year XVIII. Unless a perfect score is required, the Fables are too easy as a test for these students.

As the numerical grades of all these students are recorded, the averages of all the subjects taken by them during their college course (in the case of the high school girls during their high school course) were used as an index of their scholarship and correlated with their Point Scale and Terman averages, giving the following coefficients (r). The figures in parenthesis indicate the number of students in each class.

TABLE IV

	Sophomores	Juniors-Seniors	Coll.	St.Teachers	Jun.	Sen.	H.S.
Point Scale av. & Scholarship	.43(50)	.73(81)	.74(131)		.33	.38	.39
Point Scale av. & Psychology av.	.56	.69	.68	.85(17)			
Terman XIV–XVIII & Scholarship	.21(48)	.38(92)	.30(140)				
Terman XIV–XVIII & Psychology	.13	.18	.16				
Terman XVI–XVIII & Scholarship	.30	.55	.52				
Terman XVI–XVIII & Psychology	.46	.65	.65	.47(17)			
Scholarship & Psychology	.48	.41	.50				
Point Scale & Terman XIV–XVIII	.37(44)	.36(57)	.39(101)				
Point Scale & Terman XVI–XVIII	.37	.22	.49		.51(118)		

The most striking feature about these coefficients is that the averages of the Terman tests give much lower correlation than the Point Scale scores. This holds true also in all my investigations with younger children. The coefficients for the high school classes are low, but the number of pupils, only 39 in all, is too small to give them much significance. It is interesting, however, to note that the coefficients increase with the higher classes, thus showing the effect of the greater maturity and the larger number of courses included in the scholarship average. It is evident that the Terman tests for the years XVI and XVIII give much higher correlation with scholarship than the combined XIV, XVI, and XVIII year exercises. In the case of psychology, the average of only the first year's courses was used, so that all the classes would be rated on the same work. Again, the Point Scale correlations are higher, tho the coefficients for the Terman Adult tests (XVI-XVIII) are nearly the same as for the Point Scale.

In conclusion, the results of these tests show that the averages of these students for the Adolescent-Adult Point Scale was higher than that obtained by Senior women at Harvard and that the dis-

tribution of the scores resemble more closely that of the Harvard men, except at the lower end. The fact that these students were less mature and that the majority were Sophomores and Juniors will account for some of the low scores. These results also show that some of the tests are of doubtful value in differentiating the more from the less mature and able students. The correlations between mental and educational performance give higher coefficients for the Yerkes-Rossy Adolescent-Adult Point Scale than for the Terman tests for years XIV-XVIII, and also for the years XVI-XVIII, tho in the latter case the difference is not so great.

Dagny Sunne

Newcomb College.

THE GROWTH OF INTELLIGENCE

In connection with some recent investigations on the growth of intelligence, I have been astonished by one fact which is so strikingly substantiated by several sets of data and is of such great importance, if subsequently verified, that I venture to submit it to the readers of the Journal in advance of formal presentation in a technical article.

The investigations of Terman seemed to indicate that 16 years was the upper life age limit in the growth of intelligence of average individuals. New data seem to indicate that this age should be reduced to 13 years. The facts in support of such reduction are as follows:

1. The average mental age of drafted recruits in the U. S. Army by group tests and by mental age equivalents for group test scores as indicated by clinical examinations, proved to be approximately 13 years. This may be taken to mean that mental age growth on the average was complete in these individuals at 13 years of age or thereabouts.

2. The application of the Army Alpha group test to over 500 children in a typical average grammar school of Trenton, N. J., showed increasing median scores up to the age of 13 with no increase thereafter, although a significant number of cases in the ages up to 16 years were examined. This was true for both sexes.

3. The application of the Army Alpha group test to approximately 500 delinquent boys in the State Home for Boys at Jamesburg, N. J., showed increasing medium scores for each age group up to 13 without increasing thereafter, although a sufficient number of cases was available up to 16 years.

4. The study of continuous mental age examinations of feeble-minded from the Training School at Vineland, covering at least four successive annual examinations, indicates that mental age continues to develop in all grades of feeble-mindedness up to the life age limit of 13 years. Out of 250 cases to be studied only 30

have so far been statistically examined, but the conclusion seems legitimate from an inspection of the uncalculated results. If this conclusion be substantiated, then it follows that the point of final arrest in the mental development of the feeble-minded is, after all, not a function of retardation or of mental grade, but is conditioned by life age.

These conclusions are in accord with the very early investigations conducted by Gilbert at Yale about 25 years ago. They are also in accord with the most recent investigation on brain development. I am informed by Dr. S. D. Porteous, Director of Research at the Training School at Vineland, that his extensive investigation of cranial and brain development indicates that at 13 years brain development is practically complete in gross volume although a certain amount of cortical elaboration goes on after that age. This elaboration, in his view, is more the result or basis of intellectual maturity or experience rather than of brain power as such; that is, the development after 13 years may be qualitative rather than quantitative, although, of course, in this case quality also is quantity.

These several lines of investigation agree so consistently in their combined conclusion that I submit it as a new working hypothesis in psychological investigation as applied to education. It is at present only an hypothesis indicated by preliminary investigation and is not in any sense a final conclusion. Its import for educational theory and practice is obvious. It implies that education for the masses is available on an intellectual basis only through the grammar grades. High-school and college work from this point of view is available only for those above the 13-year median intelligence, probably the upper quartile of all ages above 13. If substantiated, this 13-year level will necessitate a revision of all previous work based on intelligence quotients which were calculated on the 16-year basis. It will also revise our concepts in relation to gifted individuals in ages after 13.

Further communication on this topic both theoretical and as a result of experiment is highly desirable. It is to be hoped that this preliminary communication will stimulate reaction in other psychologists.

EDGAR A. DOLL.
Psychologist, Department Institutions and Agencies, Trenton, N. J.

ABSTRACTS AND REVIEWS

Abraham A. Roback. *The Interference of Will-Impulses, With Applications to Pedagogy, Ethics, and Practical Efficiency.* Psychological Monographs, Vol. 25, No. 5, Whole Number III, 1918. Pp. 158.

The experimentation of this study consisted in arranging a series of reactions to a series of rapidly presented stimuli. These stimuli were series of red crosses and green dots presented to the subjects in a succession that sometimes began at a relatively slow and increased to a rapid rate and sometimes at a rapid and decreased to a slow rate. Each presentation was in two periods of about three minutes each, separated by a rest interval. The motor responses consisted of all sorts of movements of two telegraph keys with varied combinations between the movements demanded and the crosses and dots that served as stimuli. The general plan was to study the phenomena of interference which occurred when different and somewhat conflicting responses had to be made in rapid succession.

Passing over the experimental data themselves, we may refer to a number of inferences and applications that Mr. Roback enters upon toward the close of his monograph. Among these the following are of interest:

1. In a conflict of two interests, the line of least resistance is followed. If a man chooses, then, to do right, it is because that desire is stronger in him than the desire to do wrong.
2. Of two tasks of unequal difficulty, it is better to begin with the more difficult one if both are relatively easy, but with the less difficult one if both are relatively difficult.
3. Because in the experimental work it was found that after an inhibitory break the situation is usually regained by starting on the easier of the two movements, it is inferred that there should be a review of previous work before a new line of thought is taken up.
4. Mistakes occur as frequently when a task is too easy as when a task is too difficult. A certain moderate amount of difficulty should attend each task.
5. The amount of ease or difficulty experienced in attacking a task depends partly on the attitude with which it is confronted. For instance, a very difficult task may be accomplished with more than ordinary success when the person is 'set' for an artificial task even more exacting.
6. There seem to be differences in motor type which ought to be recognized by teachers in dealing with pupils. In this experiment the subjects fall into two groups—the 'undulators,' who emphasize certain tasks and leave gaps in the performance of others, and the 'equalizers,' who distribute their efforts over all tasks but may fail to complete any of them.

7. When confusion or difficulty appears, the more specific determining tendencies disappear before the more general ones do. For instance, a subject may continue to tap the key when he no longer taps it in the precise manner prescribed. From this it is inferred that the larger general principles received in school instruction will persist when specific details are lost, so that the closest attention and greatest stress should be laid on the more specific aspects of instruction.

8. It may be possible to determine the degree of a person's susceptibility to interference by a series of tests and these determinations might conceivably be significant for industrial efficiency.

These applications are surely interesting. One may be disposed to grant their correctness as educational maxims and yet question whether they can be legitimately inferred from the experiments made by Mr. Roback. They strike one as being examples of 'drawing a long bow.'

HELEN DAVIS.

Carnegie Institute of Technology.

LOUIS C. KARPINSKI. *Origines et Developpement de l'Algebre.* Scientia, Volume 26, pages 89-101.

The historical development of our system of elementary algebra has considerable significance for the student of education. Algebra is not a game, devised by some single individual and foisted off on an unsuspecting world as a subject of instruction for the youth of the land; rather algebra is the almost inevitable product of intelligence (reason) reacting upon the world in which it finds itself.

The ancient Egyptians as early as 2000 B. C. began the study of first degree equations and even types of quadratics, with some algebraic symbolism; they also discussed somewhat intensively arithmetical and geometrical series, powers of numbers, formulas for areas and volumes, and other algebraical notions which are still the subject of instruction. Independently the Babylonian civilization made studies along these various lines, excepting only the equations.

The direction of progress of Greek studies along algebraical lines was determined largely by the preceding Egyptian work, but the Greek treatment of many of these notions was geometrical; the Greek contribution to progress was primarily the rigorously logical procedure.

The Hindu civilization produced similar algebraical material, but emphasized rather algebraical notations and numerical computations.

To the Arabs we are indebted not only for the first systematic treatises on algebra, but also for systematic discussions of the Hindu-Arabic arithmetic which we use, and of trigonometry. The Arabs are by no means simple intermediaries without significant contributions of their own. They combined the logical Greek material

with the numerical and strictly algebraical ideas of the Hindus and they systematized the resultant whole.

The revival of learning in Europe began with the study of Arabic works of science and philosophy. Through translation of Arabic works, the algebra and the new arithmetic were spread broadcast through medieval Europe. The first fruits of this revival are found in Italy, in the solution of the cubic and the biquadratic. The further development continued in France, particularly with Viete, to whom we are indebted for a general literal symbolism, and with Descartes and Fermat (both 17th century) to whom we owe the graphical representation of algebraical equations. Later developments are the products of mathematicians of all the world.

Our elementary instruction in algebra has undoubtedly failed, frequently, to bring out the universality of the appeal and of the application of this subject, but this defect demands not eradication of the algebra but modification of our treatment of it in the school.

Abstract by the AUTHOR.

University of Michigan.

The Journal of Educational Psychology

INCLUDING EXPERIMENTAL PEDAGOGY, CHILD PHYSIOLOGY
AND HYGIENE, AND EDUCATIONAL STATISTICS

Published monthly except June to August by
WARWICK & YORK, INC.
HERSHEY, PA. BALTIMORE, MD.

Manuscripts for publication, books or other material for review, and news items of interest to educational psychologists should be addressed to The Managing Editor, J. Carleton Bell, 1032A Sterling Place, Brooklyn, N. Y.

NOTICE TO SUBSCRIBERS. Notify the publishers, Warwick & York, Inc., of any change of address; as copies lost in transit because of subscriber's removal to another address will not be duplicated

THE JOURNAL is issuing a series of Monograph Supplements for the publicaiion of more extensive reseaıches in educational psychology. For details prospective contributors may apply to H. E. Buchholz, 10 East Center St., Baltimore, Md.

The price of the JOURNAL is $3 00 pei year of nine monthly issues, payable in advance. Single issues are 60 cents each Back volumes are $5.00 in original wrappers and $6.75 bound in library buckram. Foreign postage is 20 cents extra for Canada and 30 cents for other countries in the Postal Union.

Subscriptions and correspondence concerning business matters should be sent to Warwick & York, Inc.

VOLUME X DECEMBER NUMBER 10

EDITORIAL

A DECADE OF EDUCATIONAL PROGRESS

With the present number the JOURNAL OF EDUCATIONAL PSYCHOLOGY completes its tenth year. The decade has been characterized by intense and diversified activity in the scientific study of education. A brief glance at some of its outstanding features will be in place here. At the beginning of the decade the old "child study" movement had reached its zenith and its methods of "questionnaire" investigation had been severely criticized. It had, however, aroused a widespread popular interest in the study of educational processes, and had contributed in no small degree to the establishment of teaching positions in education in the colleges and universities of the country. Meumann had published his *Vorlesungen zur Einfuehrung in die Experimentelle Paedagogik* in 1907, and through this work and his *Zeitschrift* maintained a dominating position in the scientific study of education throughout the world. There was much talk of the applications of laboratory psychology to education, but the psy-

chology contemplated was of the formal, technical, Wundtian type, and the applications were neither numerous nor helpful. The Binet tests, through Goddards missionary efforts, were beginning to be talked of here and there, Rice and Stone had attacked the problem of arithmetical abilities, Rice and Cornman had made studies in spelling, and Thorndike was gathering materials for his handwriting scale,—the first scale to be offered for measuring attainments in a school subject. At this juncture the JOURNAL OF EDUCATIONAL PSYCHOLOGY was established, as an organ for the publication of educational researches, as a forum for the discussion of problems in educational psychology, and as a mediator between the laboratory investigator and practical school man.

A detailed survey of the activities in the scientific study of education during the decade, and of the part which this JOURNAL has played in them, would take us far beyond our present limits. In the rapidly expanding field of tests of intelligence the studies of Terman on the revision of the Binet scale, the various articles expounding the Yerkes Point Scale, the Otis Group Tests of Intelligence, and the Pressey Primary Tests are some of the more important of the contributions appearing in its columns. In the general subject of learning we have had the studies of Dearborn and others on the effects of practice, and those of Pyle, Lakenan, Lyon, and Myers on memory, and those of Starch, Hollingworth and Poffenberger on correlation and the spread of improvement. Other interesting studies in general educational psychology have been those of Foster, Whipple and Dallenbach on observation, that of Yerkes in advertising, that of Jones in telegraphy, and more recently that of Herring in the measurement of thinking. But it is in the field of the measurement of attainments in school subjects that the most conspicuous service has been rendered. Suffice it to note the articles of Pintner, Starch, Meade, Peters, Whipple, Thorndike, Kelly, and Monroe on reading, those of Winch, Pearson, Cook, Kline, Starch, and Murray on spelling, those of Starch, King and Johnson, Pintner, and Myers and Lister on handwriting, those of Winch, Starch, Brown Phillir, Hahn and Thorndike, Kirkpatrick, and Cobb on arithmetic, of Thorndike on composition, of Rapeer, and Starch on grammar, of Lackey on geography, and of Childs on drawing. Ruediger and Strayer, Witham, and Boyce have proposed scales for the measurement of merit in teachers, Henmon has described valuable Latin scales, Minnick, and Stockard and Bell have discussed tests in geometry, Bell and McCollum have reported tests in United States history, Sackett has presented a scale in ancient history, and Bell has given the results of a test in chemistry. It will be seen, therefore, that in every field of educational measurement the JOURNAL has taken a prominent part, and has contributed no small share to that vigorous development of the scientific study of education which has been the chief characteristic of the decade, and which promises to put America in the forefront of educational progress throughout the world.

J. C. BELL

NOTES AND NEWS

The November meeting of the New York Society for the Experimental Study of Education was devoted to the topic of "Problems in the Teaching of Secondary Mathematics." The chief speakers were Dr. J. H. Minnick, of the University of Pennsylvania, on "Mathematical Tests," and Dr. Truman Lee Kelley, of Teachers College, on "Purposes of High School Mathematics." Dr. Minnick made a critical survey of the various tests in algebra and geometry, and discussed their value for diagnosis, prognosis, and measurement of mathematical attainment. More extended experimentation is called for in working out new tests for specific ends, comparing present tests for their usefulness in school work, and evaluating the results from the point of view of the class-room teacher. Dr. Kelley gave a digest of an investigation to determine the values of high school mathematics in the estimation of prominent business and professional men and women. On the basis of the findings an experiment was arranged to teach first year high school mathematics as an inspirational and cultural subject to one class, while a parallel class was taught traditional algebra. Standard tests were used at the beginning and at the end of the experiment. The experimental class showed a keener interest and a broader grasp, while the control class was superior in the drill aspects of mathematics.

The New York Branch of the American Psychological Association met on Monday, November 24, with the following program: "Emotional Instability of Children," Miss Buford Johnson; "A Criticism of MacDougall's Doctrine of Interests," Mr. Alfred H. Martin; "A Study of Race Difference in Intelligence," Miss Katherine Murdock; "The Relation between Intellectuality and Morality," Miss Clara F. Chassell.

Dr. J. E. W. Wallin, who has been chairman of the Committee on Defective Children for the Missouri Children's Code Commission during the last four years, has been appointed chairman of the Committee on Mental Defectiveness for the Missouri Conference for Social Welfare, and elected president of the Department of Special Classes of the Missouri State Teachers Association.

Dr. T. R. Garth, of the department of psychology in the West Texas State Normal College, has been appointed adjunct professor of psychology at the University of Texas.

Lieutenant-Colonel C. S. Myers, university lecturer and director of the university laboratory of experimental psychology, Cambridge, England, has been elected to a fellowship in Gonville and Caius College, Cambridge.

Dr. Edward G. Boring, of Cornell University, has been appointed professor of experimental psychology and head of the psychological laboratory at Clark University to succeed the late Professor Baird.

Dr. J. E. Dashiell, assistant professor of psychology in Oberlin College, has been appointed associate professor in charge of psychology in the University of North Carolina, succeeding in that capacity the president elect, Dr. H. W. Chase.

Dr. A. E. Davies, professor of philosophy in Ohio State University, has been appointed professor of philosophy and psychology in Colorado College.

Mr. Homer W. Anderson has been appointed to a position in the department of research of the Detroit city schools.

Mr. David A. Grove has been appointed professor of psychology and education at the Western College for Women, Osford, Ohio.

Dr. Chester A. Buckner, for the past two years in charge of educational measurements in The Lincoln School, Teachers College, has been appointed professor of education in the University of Kansas, where he will be the director of the Bureau of School Service.

Dr. John M. Brewers, head of the department of education of the Southern Branch of the University of California, has accepted the position of director of the Harvard Bureau of Vocational Guidance.

Dr. Rutledge T. Wiltbank, of the University of Washingron, has been appointed assistant professor of psychology in the University of Chicago.

Dr. Walter R. Ames has been appointed assistant professor of education and psychology in the University of Montana.

Journal of Educational Psychology
Vol. 10, 1919.